

-2099-

++G +S G+IL+HGV+ E T FNGI I GA + +Q RVLMLS+KAR DANPIL
 Sbjct: 306 HHGKHSEGYILKHGVMREAAATSFNGISKIEHGATKSHGEQTERVLMLEKARGDANPIL 365

Query: 349 LIDENDVTAGHAASIGQVDPEDLYYLMRGLNOKTAEQLVIRGFLGTVIAEIPVKEVRDE 408
 LIDE+DVTAGHAAS+G++DP ++YLMRSG+++ AE+LVI GFL V+ ++P++ V++

Sbjct: 366 LIDEDDVTAGHAASVGKIDPIQMFYLMRSGISRAEERLVIHGFLAPVVGQLPIESVKER 425

Query: 409 MIAVIDTKLE 418
 ++ I+ K++

Sbjct: 426 LVEAIERKVK 435

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5783> which encodes the amino acid sequence <SEQ ID 5784>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.80 Transmembrane 387 - 403 (387 - 403)

----- Final Results -----

bacterial membrane --- Certainty=0.1319(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB15259 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 177/428 (41%), Positives = 267/428 (62%), Gaps = 15/428 (3%)

Query: 3 KEKLVAFSQAHAEPWLQERRLAALAEIAPNLELPTIERVKFHRWNLGDGT--LTENESLA 60
 +E L +FS+ H EPAWL+ RL ALE +L +P ++ K WN + +NE L+

Sbjct: 11 QEYLKSFSEKHQEPWLKNLRLQALEQAEDLPMFKPKDKTKITNWNFTINFAKHTVDNEPLS 70

Query: 61 SVPDF-----IAIGDNPKLQVGTQTVLEQLPMA--LIDKGVVFSDFYTALEEIPEVI 111
 S+ D I I + K + V L ++ L DKGV+F+D TA E +++

Sbjct: 71 SLEDLTDEVKALIDIENEDKTLVQQRDQTPAHLSSLQELKDKGVIFTDILTAAREHSDLV 130

Query: 112 EAHFGQ-ALAFDEDKLAAYHTAYFNSSAAVLYVPDHLEITTPIEAIFLQSDSDVPFNKHV 170
 E +F + + DE KL A H A N A LYVP +++ TP++A+++ +S+ FN HV

Sbjct: 131 EKYFMKDGKVKVDEHKLTAALVNGGAFLYVPKNVQVETPVQAVVHESNDTALFN-HV 189

Query: 171 LVIAGKESKFTYLERFESIGNATQKISANISVEVIAQAGSQIKFSAIDRLGPSVTITYISR 230
 L++A S TY+E + S N + NI EVI + + + A+D L VTTY++R

Sbjct: 190 LIVAEHDSSVTYVENYISTVNPKDAVF-NIIEVITGDNASVTYGAVDNLSSGVTTYVNR 248

Query: 231 RGRLE-KDANIDWALAVMNEGNVIADFDSDLIGQGSQADLKVVAASSGRQVQGDITRVIN 289
 RG +D+ I+WAL +MN+G+ I++ ++L G G+ D K V G Q + T++ +

Sbjct: 249 RGAARGRDSKIEWALGLMNDGDTISENTNLYGDTYGDTKTVVVGGRGEQTENFTTQIIH 308

Query: 290 YGQRTVGHILQHGVLERGTITFNGIGHILKDAKGADAQQESRVLMLESDQARADANPILL 349
 +G+ + G+IL+HGV+ + + FNGIG I A A+A+QESRVLMLE++AR DANPILL

Sbjct: 309 FGKASEGYILKHGVMKDSASSIFNGIGKIEHGASKANAEQESRVLMLEKARGDANPILL 368

Query: 350 IDENEVTAGHAASIGQVDPEDMYLMSRGLDQETAERLVIRGFLGAVIAEIPISVRQEI 409
 IDE++VTAGHAAS+G+VDP +YYLMRSG+ +E AERLVI GFL V+ E+PI V++++

Sbjct: 369 IDEDDVTAGHAASVGRVDPIQLYYLMRSGIPKEEAERLVIYGFAPVVNLEPIEGVKKQL 428

Query: 410 IKVLDEKL 417
 + V++ K+

Sbjct: 429 VSVIERKV 436

An alignment of the GAS and GBS proteins is shown below.

Identities = 322/420 (76%), Positives = 368/420 (86%)

Query: 1 MSKEAILNPLQAKGEPTWLQELRLKAPEKIEELPVIKRVKPHRWNLGDGTILENDYTA 60
 M+KE ++ F QA EP WLQE RL A E I LELP IERVKPHRWNLGDGT+ EN+ A

-2100-

Sbjct: 1 MTKEKLVAFSQAHAEPALWQERRLAALAEIAPNLELPTIERVKFHRWNLGDGTLTENESLA 60

Query: 61 NVPDFTELGNPKLVQIGTQTVLEQVPMELIEKGVVFTDFYSALBEEIPEVIERYFGKARP 120
+VPDF +G+NPKLVQ+GTQTVLEQ+PM LI+KGVVF+DFY+ALBEEIPEVIE +FG+A

5 Sbjct: 61 SVPDFIAIGDNPVKLVQVGTQTVLEQLPMALIDKGVVFSDFYTALEBEEIPEVIEAHFGQALA 120

Query: 121 FEEDRLAAYHTAYFNSGAVLYIPDNVEITQPIEGLFYQDSQSKVPFNKHILLIVGKNAKV 180
F+ED+LAAYHTAYFNS AVLY+PD++EIT PIE +F QDS S VPFNKH+L+I GK +K

10 Sbjct: 121 FDEDKLAAYHTAYFNSAAVLYVPDHLLEITTPIEAIFLQSDSDVPFNKHVLVIAGKESKF 180

Query: 181 SYLERFESIGDGTERTSANISVEVIAQAGSQIKFASIDRLGENVTTFISRRGRHSSDATI 240
+YLERFESIG+ T++ SANISVEVIAQAGSQIKF++IDRLG +VTT+ISRRGR DA I

15 Sbjct: 181 TYLERFESIGNATQKISANISVEVIAQAGSQIKFSAIDRLGPSVTITYISRRGRLEKDANI 240

Query: 241 DWALGVMNEGNVADFDSDLIGDGSANLKVVAASSGRQVQIDTRVTNYGCNSVGHILQ 300
DWAL VMNEGNV+ADFDSDLIG GS A+LKVVAASSGRQVQIDTRVTNYG +VGHILQ

20 Sbjct: 241 DWALAVMNEGNVIADFDSDLIGQSQADLKVVAASSGRQVQIDTRVTNYGQRTVGHILQ 300

Query: 301 HGVILERGTLTFNGIGHI+KAGKAGADAQQESRVLMLSDKARSDANPILLIDENDVTAGHA 360
HGVILERGTLTFNGIGHI+K AGKAGADAQQESRVLMLSD+AR+DANPILLIDEN+VTAGHA

25 Sbjct: 301 HGVILERGTLTFNGIGHILKDAKAGADAQQESRVLMLSDQARADANPILLIDENEVTAGHA 360

Query: 361 ASIGQVDPEDLYYLMRGLNQKTAEQLVIRGFLGTVIAEIPVKEVRDEMAVIDTKLEKR 420
ASIGQVDPED+YYLMRGL+Q+TAE+LVIRGFLG VIAEIP+ VR E+I V+D KL R

30 Sbjct: 361 ASIGQVDPEDMYLMRGLDQETAERLVIRGFLGAVIAEIPPSVRQEIIVLDEKLLNR 420

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1861

30 A DNA sequence (GBSx1968) was identified in *Sagalactiae* <SEQ ID 5785> which encodes the amino acid sequence <SEQ ID 5786>. This protein is predicted to be ABC transporter, ATP-binding protein, Ycf16 family. Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2253(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15260 GB:Z99120 similar to ABC transporter (ATP-binding protein) [Bacillus subtilis]
Identities = 180/250 (72%), Positives = 212/250 (84%)

45 Query: 2 SVLEIKNLHVSIEDKEILKGLNLTKTGEIAAIMGPNGTGKSTLSAAIMGNPNVEVTAGE 61
S L IK+LHV IE KEILKG+NL +K GE A+MGPNGTGKSTLSAAIMG+P YEVT G

Sbjct: 4 STLTIKDLHVEIEGKEILKGVNLEIKGGEFHAVMGPNGTGKSTLSAAIMGHKPYEVTKGS 63

50 Query: 62 ILFDGEDILELEVDERARLGLFLAMQYPSEVPGITNAEFIRAAMNAGADDDKISIRQFI 121
I DG+D+LE+EVDERA+ GLFLAMQYPSE+ G+TNA+F+R+A+NA + + D+IS+ +FI

Sbjct: 64 ITLDGKDVLEMEVDERAQAGLFLAMQYPSEISGVTNADFLRSAINARREEGDEISLMKFI 123

55 Query: 122 TKLDEKMELLGMKEEMAERYLNEGFSGGEEKRNEILQLLMLEPKFALLDEIDSGLDIDAL 181
K+DE ME L M EMA+RYLNEGFSGGEEKRNEILQL+M+EPK A+LDEIDSGLDIDAL

Sbjct: 124 RKMDENNEFLEMDPEMAQRYLNEGFSGGEEKRNEILQLMMIEPKIALDEIDSGLDIDAL 183

Query: 182 KVVSKGVNEMRGEFGGAMIITHYQRLNLYITPDVHVMMDGKVVLSGGPELAVRLEKEGY 241
KVVSKG+N+MR E FG ++ITHYQRLNLYITPD VHVMM G+VV SGG ELA RLE EGY

60 Sbjct: 184 KVVSKGINKMRSENFGLMITHYQRLNLYITPDVHVMMQGRVVKSGGAELAQRLAEGY 243

-2101-

Query: 242 AQIAEELGLE 251
 I +ELG+E
 Sbjct: 244 DWIKQELGIE 253

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5787> which encodes the amino acid sequence <SEQ ID 5788>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2417(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 225/255 (88%), Positives = 241/255 (94%)

Query: 1 MSVLEIKNLHVSIEDKEILKGINLTLKTGEIAAIMGPNGTGKSTLSAAIMGPNPYEVTAG 60
 MS+LEI NLHVSTI KEILKG+NLTLKTGE+AAIMGPNGTGKSTLSAAIMGPNPYEVT G
 20 Sbjct: 1 MSILEINNHLHVSIEGKEILKGVNLTTLKTGEVAAIMGPNGTGKSTLSAAIMGPNPYEVTQG 60
 Query: 61 EILFDGEDILELEVDERARLGLFLAMQYPSEVPGITNAEFIRAAMNAGKADDDKISIRQF 120
 +IL DG +IL+LEVDERARLGLFLAMQYPSE+PGITNAEF+RAAMNAGKAD+DKIS+R F
 25 Sbjct: 61 QILLDGVNILDLEVDERARLGLFLAMQYPSEIPGITNAEFMRAAMNAGKADEDKISVRDF 120
 Query: 121 ITKLEDEKMELLMKEEMAERYLNEGFSGGEEKRNEILQLLMLEPKFALLDEIDSGLDIDA 180
 ITKLEDEKM LLGMKEEMAERYLNEGFSGGEEKRNEILQLLMLEPKFALLDEIDSGLDIDA
 Sbjct: 121 ITKLEDEKMALLGMKEEMAERYLNEGFSGGEEKRNEILQLLMLEPKFALLDEIDSGLDIDA 180
 30 Query: 181 LKVVSCKGVNEMRGEGFGAMITHYQRLNLYITPDKVHVMMDGKVVLSGGPELAVRLEKEG 240
 LKVVSCKGVNEMRG+ FGAMITHYQRLNLYITPD VHVMMDG++VLSG LA RLEKEG
 Sbjct: 181 LKVVSCKGVNEMRGKDFGAMITHYQRLNLYITPDLVHVMMDGRIVLSGDAALATRLEKEG 240
 35 Query: 241 YAQIAEELGLEYLEYKEE 255
 YA IA++LG+EYKEE
 Sbjct: 241 YAGIAQDLGLEYLEYKEE 255

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 1862

A DNA sequence (GBSx1969) was identified in *S.agalactiae* <SEQ ID 5789> which encodes the amino acid sequence <SEQ ID 5790>. This protein is predicted to be RgpG (rfe). Analysis of this protein sequence reveals the following:

Possible site: 40
 45 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -12.10 Transmembrane 312 - 328 (308 - 336)
 INTEGRAL Likelihood = -10.03 Transmembrane 15 - 31 (6 - 41)
 INTEGRAL Likelihood = -9.82 Transmembrane 205 - 221 (197 - 226)
 50 INTEGRAL Likelihood = -8.60 Transmembrane 335 - 351 (329 - 358)
 INTEGRAL Likelihood = -7.48 Transmembrane 257 - 273 (255 - 281)
 INTEGRAL Likelihood = -5.52 Transmembrane 60 - 76 (56 - 79)
 INTEGRAL Likelihood = -5.31 Transmembrane 151 - 167 (148 - 171)
 INTEGRAL Likelihood = -4.88 Transmembrane 91 - 107 (90 - 108)
 INTEGRAL Likelihood = -4.78 Transmembrane 184 - 200 (177 - 203)
 55 INTEGRAL Likelihood = -3.13 Transmembrane 119 - 135 (119 - 135)
 INTEGRAL Likelihood = -2.97 Transmembrane 229 - 245 (229 - 250)

----- Final Results -----
 bacterial membrane --- Certainty=0.5840(Affirmative) < succ>

-2102-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8919> which encodes amino acid sequence <SEQ ID 8920> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10

McG: Discrim Score: 5.18

GvH: Signal Score (-7.5): -6.19

Possible site: 15

>>> Seems to have an uncleavable N-term signal seq

ALOM program count: 9 value: -12.10 threshold: 0.0

INTEGRAL Likelihood = -12.10 Transmembrane 239 - 255 (235 - 263)

INTEGRAL Likelihood = -9.82 Transmembrane 132 - 148 (124 - 153)

INTEGRAL Likelihood = -8.60 Transmembrane 262 - 278 (256 - 285)

INTEGRAL Likelihood = -7.48 Transmembrane 184 - 200 (182 - 208)

INTEGRAL Likelihood = -5.31 Transmembrane 78 - 94 (75 - 98)

INTEGRAL Likelihood = -4.88 Transmembrane 18 - 34 (17 - 35)

INTEGRAL Likelihood = -4.78 Transmembrane 111 - 127 (104 - 130)

INTEGRAL Likelihood = -3.13 Transmembrane 46 - 62 (46 - 62)

INTEGRAL Likelihood = -2.97 Transmembrane 156 - 172 (156 - 177)

PERIPHERAL Likelihood = 12.63 284

modified ALOM score: 2.92

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.5840(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA82114 GB:AB022909 RgpG [Streptococcus mutans]

Identities = 266/382 (69%), Positives = 317/382 (82%)

Query: 10 TIEYIFVLIGAFLLSIILTPIIIRVISLKVGAVDKPNARRINKVPMPSGGGLAIFLSFVVT 69

T++++ VLI L S++LTP++R +L+VGAVD PNARRINKVPMPS+GGLAI +SFV+

Sbjct: 7 TLKFVLVLIIATLLTSLVLTPLVRFALRVGAVDNPARRINKVPMPSAGGLAIIISFVIA 66

Query: 70 TLFFMPMAASRHFIEVSYPHYILPVIIGGLVVTTFGIDDIFELRPYKMLGIIIAAII 129

TL MPM SYF YILPV++G LV+ TGFIDD++EL P+ K LGI++ A+II

Sbjct: 67 TLALPMILKTQIGGKSYFEYILPVVLGALVIALTGFIDDVYELSPKIKFLGILLGAVII 126

Query: 130 WKFTFRFDSFKIPIGGPLLEFGPILTFFLTIVLWIIISITNAINLIDGLDGLVSGVSIISL 189

W FT FRFDSFKIP GGP+L F P L+FFLT+LW+++ITNA+NLIDGLDGLVSGVS+ISL

Sbjct: 127 WIFTFRFRDSFKIPFGGPMLEHFNFLSFFLTILVWVAITNAVNLIDGLDGLVSGVSMISL 186

Query: 190 ATMAVVSYPFLPKIDFPLTLTIVILIASIVGFFPYNYHPAIIYLG DAGALFIGFMIGVLS 249

TM +VSYPFL D FLTLTI +LI +I GFFPYNYHPAIIYLG DAGALFIGFMISVLS

Sbjct: 187 TTMGLVSYPFLYDITDIFLTLTIFVLIFAIAGFFPYNYHPAIIYLGDTGALFIGFMISVLS 246

Query: 250 LQGLKNSTAVAVITPVIILGVPILDTAVAI VRRKLSGKKISEADKMHLHRLLSMGFTHR 309

LQGLKN+TAVAV+TP+I+LGVPI+DT VAI+RR LSG+K EAD MHLHRL+MGFTHR

Sbjct: 247 LQGLKNATAVAVTPITVLGVPIVDTTVAIIRRTLSGQKFYEADNMHLHRLLAMGFTHR 306

Query: 310 GAVLVVYGIATIFSLIALLLNVSSRIGGIFLLALLAMEIFIEGLNIWGENRTPLFNLL 369

GAVLVVYGIA+ FSL++LLLVSSR+GGI L++ + A+EIFIEGL IWG RTPLF LL

Sbjct: 307 GAVLVVYGIAMFFSLVSLLLNVSSRLGGIILLMIGVAFALIEFIEGLEIWGPKRTPLFRLL 366

Query: 370 KFIGNSDYRQSVIAKYSDKHQK 391

FIGNSDYRQ V+AKY K +K

Sbjct: 367 AFIGNSDYRQEVAKYRRKKKK 388

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5791> which encodes the amino acid sequence <SEQ ID 5792>. Analysis of this protein sequence reveals the following:

-2103-

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq

5	INTEGRAL	Likelihood = -8.28	Transmembrane	9 - 25 (1 - 33)
	INTEGRAL	Likelihood = -8.17	Transmembrane	201 - 217 (198 - 221)
	INTEGRAL	Likelihood = -7.64	Transmembrane	308 - 324 (305 - 329)
	INTEGRAL	Likelihood = -7.17	Transmembrane	55 - 71 (51 - 74)
	INTEGRAL	Likelihood = -7.06	Transmembrane	145 - 161 (138 - 170)
	INTEGRAL	Likelihood = -6.58	Transmembrane	260 - 276 (251 - 278)
10	INTEGRAL	Likelihood = -6.21	Transmembrane	180 - 196 (172 - 198)
	INTEGRAL	Likelihood = -5.95	Transmembrane	331 - 347 (330 - 353)
	INTEGRAL	Likelihood = -5.68	Transmembrane	87 - 103 (82 - 104)
	INTEGRAL	Likelihood = -3.93	Transmembrane	113 - 129 (112 - 133)
	INTEGRAL	Likelihood = -2.60	Transmembrane	233 - 249 (232 - 250)

15 ----- Final Results -----
 bacterial membrane --- Certainty=0.4312(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20 The protein has homology with the following sequences in the databases:

>GP:BAA82114 GB:AB022909 RgpG [Streptococcus mutans]
 Identities = 289/381 (75%), Positives = 334/381 (86%), Gaps = 1/381 (0%)

25	Query: 5	TIDYVLVLIGALLMSLFLTPLVRFALFRVGAVDNPVNARRVNVKVPMPSTSGGLAIFMSFLVA 64
		T+ +VLVLI LL SL LTPLVRF A RVGAVDNPVNARR+NKVPMP++GGLAI +SF++A
	Sbjct: 7	TLKFVLVLIIATLLTSLVLTPLVRFALRVGAVDNPVNARRINKVPMPSSAGGLAIIISFVIA 66
	Query: 65	SLGLIPIASKGAMFFGQTYFSYILPVVIGATVITLTGFLDDLYELSPKIKMFGLIGAVI 124
		+L L+P+ K G++YF YILPVV+GA VI LTGF+DD+YELSPK+K GIL+GAVI
30	Sbjct: 67	TLALPMLIK-TQIGCKSYFEYILPVVLGALVIALTGFIDVYELSPKIKFLGILIGAVI 125
	Query: 125	VWAFDTDFKFSFKIPFGGPLLVFPGPFLTLFLTLVWIVSITNAINLIDGLDGLVSGVSIIS 184
		+W FTFD+FDSEFKIPFGGP+L F PFL+ FLT+LW+V+ITNA+NLIDGLDGLVSGVSIIS
35	Sbjct: 126	IWIFTDFRDSFKIPFGGPMHFNPFLSFFLTILWVVAITNAVNLIIDGLDGLVSGVSMIS 185
	Query: 185	LVTMAIVSYFFLPQDFLTLTLVLISAIAGFFFPYNYHPAMIYLGDTGALFIGFMIGVL 244
		L TM +VSFFL D FLTLTI VLI AIAGFFFPYNYHPA+IYLGDTGALFIGFMISVL
	Sbjct: 186	LTMGLVSYFFLYDTDFLTLTLFVLI FAIAGFFFPYNYHPAIIYLGDTGALFIGFMISVL 245
40	Query: 245	SLQGLKNSTAVAVVTPVILGVPIMDTIVAIIRRLSGQKFYEYEDKMHLLHRRLLSMGFTH 304
		SLQGLKN+TAVAVVTP+I+LGVPI+DT VAIIR+LSGQKFYE D MHLHRRLL+MGFTH
	Sbjct: 246	SLQGLKNATAVAVVTPIIIVLGVPIVDTTVAIIRRTLSGQKFYEADNMHLHRRLLAMGFTH 305
	Query: 305	RGAVLVVYGITMLFSLISLLNVSSRIGGVLLMLGLLFGLEVFIEGLEIWGEKRTPLFNL 364
45		RGAVLVVYGI M FSL+SLLLNVSSR+GG+LLM+G+ F LE+FIEGLEIWG KRTPLF L
	Sbjct: 306	RGAVLVVYGIAMFFSLVSLLLNVSSRLGGILLMIGVAFALFIEGLEIWGPKRTPLFRL 365
	Query: 365	LKFIGNSDYRQAMLLKWKKEK 385
		L FIGNSDYRQ ++ K++ KK
50	Sbjct: 366	LAFIGNSDYRQEVVAKYRRKK 386

An alignment of the GAS and GBS proteins is shown below.

Identities = 282/384 (73%), Positives = 334/384 (86%), Gaps = 1/384 (0%)

55	Query: 6	MIPFTIEYIFVLIGAFLLSIILTPRIIRVISLKVGAVDKPNARRINKVPMPSSGGLAIFLS 65
		M FTI+Y+ VLIGA L+S+ LTP++R ++ +VGAVD PNARR+NKVPMP+SGGLAIF+S
	Sbjct: 1	MFSFTIDYVLVLIGALLMSLFLTPLVRFALFRVGAVDNPVNARRVNVKVPMPSTSGGLAIFMS 60
	Query: 66	FVVTTLFFMPMAAS-RHFIEVSFYHYILPVIIIGGLVVTTTGFIDDFELRPRYKMLGIII 124
60		F+V +L +P+A+ F +YF YILPV+IG V+T TGF+DD++EL P+ KM GI+I
	Sbjct: 61	FLVASLGLIPIASKGAMFFGQTYFSYILPVVIGATVITLTGFLDDLYELSPKIKMFGLII 120
	Query: 125	AAIIIWKFTHFRDSFKIPFGGPLLVEFGPILTFFLTLVWIIISITNAINLIDGLDGLVSGV 184
		A+I+W FT F+FDSEFKIP GGPLL FGP LT FLTLVWI+SITNAINLIDGLDGLVSGV
65	Sbjct: 121	GAVIVWAFDTDFKFSFKIPFGGPLLVFPGPFLTLFLTLVWIVSITNAINLIDGLDGLVSGV 180

-2104-

Query: 185 SIISLATMAVVSFFLPKIDFFLTITIVILIASIVGFFPYNYHPAIIYLGAGALFIGFM 244
 SIISL TMA+VSFFFLP+ DFFLTITI++LI++I GFFPYNYHPA+IYLGD GALFIGFM
 Sbjct: 181 SIISLVITMAIVSYFFLPQKDFFLTITILVLISATAGFFPYNYHPAMIYLGDTGALFIGFM 240

5 Query: 245 IGVLSLQGLKNSTAVAVITPVIILGVPIIDTAVAIIVRRKLSGKKISEADKMHLHHRLLSM 304
 IGVLSLQGLKNSTAVAV+TPVIILGVPI+DT VAI+RR LSG+K E DKMHLHHRLLSM
 Sbjct: 241 IGVLSLQGLKNSTAVAVITPVIILGVPIIMDTTIVAIIRSLSGQKFYEPDKMHLHHRLLSM 300

10 Query: 305 GFTHRGAVLVVYGIATIFSLIALLLNVSSRIGGIFLLALLAMEIFIEGLNIWGENRTP 364
 GFTHRGAVLVVYGI ++FSLI+LLLNVSSRIGG+ L+L LL +E+FIEGL IWGE RTP
 Sbjct: 301 GFTHRGAVLVVYGITMLFSLISLLNVSSRIGGVLLMLGLLFGLEVFIEGLEIWGEK RTP 360

Query: 365 LFNLLKFIGNSDYRQSVIAKYSK 388
 LFNLLKFIGNSDYRQ+++ K+ +K
 15 Sbjct: 361 LFNLLKFIGNSDYRQAMLLKWKEK 384

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1863

20 A DNA sequence (GBSx1970) was identified in *S. galactiae* <SEQ ID 5793> which encodes the amino acid sequence <SEQ ID 5794>. This protein is predicted to be negative regulator of genetic competence. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3460(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30

A related GBS nucleic acid sequence <SEQ ID 9483> which encodes amino acid sequence <SEQ ID 9484> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA82113 GB:AB022909 negative regulator of genetic competence
 [Streptococcus mutans]
 Identities = 168/248 (67%), Positives = 205/248 (81%), Gaps = 9/248 (3%)

35 Query: 1 MEMKQISETTLKITISMEDLEDKDFLIPQEKTEEFFYSVMDELDLPENFKNSGML 60
 MEMKQISETTLKITISMEDLE+RGMEKDFLIPQEKTEEFFY+VMDELDLPENFK SGML
 40 Sbjct: 1 MEMKQISETTLKITISMEDLEERGMEKDFLIPQEKTEEFFYTVMDELDLPENFKGSGML 60

Query: 61 SFRVTPKKDRIDVFVTKSELSKDLNLEELADLGDISKMSPEDFFKTLQSMLEKGD TDAH 120
 SFRVTP+ DRIDVFVTKSE++K+LNLE+L+D DISKMSPEDFF TLE++M EKGD A
 45 Sbjct: 61 SFRVTPRNDRIDVFVTKSEINKNLNLEDLSDFDDISKMSPEFFNTLEETMREKGDAAAL 120

Query: 121 AKLAEIENMMDKATQEVVEENVSEEQPEKEVETIGYVHYVDFDNIEAVVRFSQTIDFPI 180
 KLAIE ++ TQ+ E+ ++E+ + YVH+V DF NI+ V+ F++T+D+ +
 50 Sbjct: 121 DKLAIEIKREEEKTQQ--EKGETKEKRD-----YVHFVLD FPNIQQVISFAKTVDYDV 171

Query: 181 EASELYKNGKGYHMTILLDLENQPSYFANLMYARMLEHANVGTKTRAYLKEHSIQLIHDD 240
 EASEL+K YHMT+LL+LE++P Y+A+LM+ARMLEHA GTKTRAYL EH +QLI D
 55 Sbjct: 172 EASELFKESDAYHMTVLLNLEDKPDYADLMFARMLEHAGRGTKTRAYLLEHGVQLIKAD 231

Query: 241 AISKLQMI 248
 A+ +LQMI
 Sbjct: 232 ALQELQMI 239

-2105-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5795> which encodes the amino acid sequence <SEQ ID 5796>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3307(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 171/253 (67%), Positives = 209/253 (82%), Gaps = 2/253 (0%)

Query: 1 MEMKQISETTLKITISMEDLEDRGMELKDFLIPQEKTEEFFYSVMDLDPENFKNSGML 60

MEMKQISETTLKITISM+DLE+RGMELKDFLIPQEKTEEFFYSVMDLDP+NFK+SGML

Sbjct: 3 MEMKQISETTLKITISMDDLEERGMEKDFLIPQEKTEEFFYSVMDLDPDNFKDSGML 62

Query: 61 SFRVTPKKDRIDVFVTKSELSKDLNLEELADLGDISKSPEDFFKTLQSMLEKGDTHDAH 120

SFRVTP+KDR+DVFTKSE++KD+NLE+LA+ GD+S+M+PEDFFK+LEQSM EKGDAH

Sbjct: 63 SFRVTPRKDRLDVFVTKSEINKDINLEDLAIEFGDMSQMTPEFFKSLEQSMREKGDVKAH 122

Query: 121 AKLAETENMDKATQEVV--EENVSEEQPEKEVETIGYVHYVDFDNIEAVVRFSQTIDF 178

KL+IE+M+ + + + ++ E E + YVHYV DF I V F++TIDF

Sbjct: 123 EKLEKIEEIMEDVVEATLANQSEAADPSTNHSEPLDYVHYVLDFTITEAVAFKTIIDF 182

Query: 179 PIEASELYKNGKGYHMTILLDDLENQPSYFANLMYARMLEHANVGTKTRAYLKEHSIQLIH 238

IEASELYK YHMTILLD++ QPSYFAN+MYAR++EHAN G+KTRAYL+EH +QL+

Sbjct: 183 SIEASELYKGSNCYHMTILLDVQQQPSYFANVMYARLIEHANPGSKTRAYLQEHGLQLML 242

Query: 239 DDAISKLMQMIEMG 251

D A+ +LQ IE+G

Sbjct: 243 DGAVEQLQKIELG 255

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1864

A DNA sequence (GBSx1971) was identified in *S.agalactiae* <SEQ ID 5797> which encodes the amino acid sequence <SEQ ID 5798>. This protein is predicted to be BacA (bacA). Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -9.02 Transmembrane 115 - 131 (111 - 135)

INTEGRAL Likelihood = -8.97 Transmembrane 227 - 243 (219 - 247)

INTEGRAL Likelihood = -7.86 Transmembrane 48 - 64 (44 - 69)

INTEGRAL Likelihood = -7.27 Transmembrane 263 - 279 (260 - 279)

INTEGRAL Likelihood = -7.22 Transmembrane 87 - 103 (85 - 107)

INTEGRAL Likelihood = -3.50 Transmembrane 2 - 18 (1 - 19)

----- Final Results -----

bacterial membrane --- Certainty=0.4609(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD50462 GB:AF169967 BacA [Flavobacterium johnsoniae]

Identities = 101/275 (36%), Positives = 165/275 (59%), Gaps = 22/275 (8%)

Query: 7 LKALFLGVVEGVTEWLPVSSTGHLILVQEFMKLNQSKSFVEMFNIVQLGAIMAVIVIYF 66

-2106-

L+A+ L V+EG+TE+LPVSSTGH+I+ F + + F ++F IVIQLGAI++V+V+YF
 Sbjct: 4 LQAIVLAVIEGITEFLPVSSSTGHMIIASSFFGIAH-EDFTKLFTIVIQLGAILSVVLYF 62
 Query: 67 KRLNPFQPGKSAREIRLTWQLWLKVVIAICIPSILIALPFDNWFEAHFNFIPIAIALIFY 126
 KR FQ T + K+++A IP++++ L ++ + +A++L+
 Sbjct: 63 KRF--FQ-----TLDYFYLKLVAFIPAVVLGLLLSDFIDGLENPVTVAVSLIG 110
 Query: 127 GFVFI---WVEKRNAHLKPQVTELASMSYKTAFLIGCFQVLSIVPGTSRSGATILGAI 182
 G + + W NA Q ++Y A IG FQ +++PG SRSGA+I+G +
 Sbjct: 111 GLILLKVDEWFNNPNAETSQ-----KITYLQALKIGLFQCIAMIPGVSRSGASIVGGMS 165
 Query: 183 IGTSRSAADFTFFLAIPTMGYSGLKAVKYFLDGNVLSLDQSLILLVASLTAFVVSILYV 242
 SR+ AA+F+FFLA+PTM G + K Y+ G LS DQ IL++ ++ AF+V+L
 Sbjct: 166 QKLSRTTAAEFSSFLAVPTMLGATVKKCYDYKAGFELSHDQVNILIIGNVAVFIVALLA 225
 Query: 243 IRFLTIDYVKRHDFITFGKYRIVLGSLLILYLVVH 277
 I+ ++ ++ F +FG YRI+ G +L+L +H
 Sbjct: 226 IKTFLTKNGFKVFGYRIIAGIILLIHHFIH 260

- 20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5799> which encodes the amino acid sequence <SEQ ID 5800>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 25 INTEGRAL Likelihood = -11.30 Transmembrane 225 - 241 (219 - 247)
 INTEGRAL Likelihood = -9.24 Transmembrane 115 - 131 (109 - 135)
 INTEGRAL Likelihood = -7.64 Transmembrane 48 - 64 (44 - 69)
 INTEGRAL Likelihood = -7.43 Transmembrane 87 - 103 (85 - 108)
 INTEGRAL Likelihood = -5.15 Transmembrane 263 - 279 (262 - 279)
 30 INTEGRAL Likelihood = -3.82 Transmembrane 2 - 18 (1 - 19)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAD50462 GB:AF169967 BacA [Flavobacterium johnsoniae]
 Identities = 102/269 (37%), Positives = 169/269 (61%), Gaps = 14/269 (5%)
 40 Query: 7 LKAIFFGIIEGITEWLPVSSSTGHLILVQEFIRLNQDKAFIEMFNIVIQLGAIIVMLIYF 66
 L+AI +IEGITE+LPVSSTGH+I+ F + + F ++F IVIQLGAI++V+++YF
 Sbjct: 4 LQAIVLAVIEGITEFLPVSSSTGHMIIASSFFGIAHED-FTKLFTIVIQLGAILSVVLYF 62
 Query: 67 ERLNPFQPGKTAREVQLTWQLWLKVVIAICIPSILIAVPLDNWFEAHFYFMVPIAIALIVY 126
 +R FQ T + K+++A IP++++ + L ++ + V +A++L++
 45 Sbjct: 63 KRF--FQ-----TLDYFYLKLVAFIPAVVLGLLLSDFIDGLENPVTVAVSLIG 110
 Query: 127 GIAFIWIEKRNAQQEPVTELARMSYKTAFFIGCFQVLSIVPGTSRSGATILGAILGTS 186
 G+ + +++ A T +++Y A IG FQ +++PG SRSGA+I+G + S
 50 Sbjct: 111 GLILLKVDEWFNNPNAETS-QKITYLQALKIGLFQCIAMIPGVSRSGASIVGGMSQKLS 169
 Query: 187 RTVAADFTFFLAIPTMGYSGLKAVKFFLDGHHLDFAQVLILLVASLTAFVVSLLAIRFL 246
 RT AA+F+FFLA+PTM G + K ++ G L QV IL++ ++ AF+V+LLAI+
 55 Sbjct: 170 RTTAAEFSSFLAVPTMLGATVKKCYDYKAGFELSHDQVNILIIGNVAVFIVALLAIKTF 229
 Query: 247 TDYVKKHDFITFGKYRIVLGSLLLIYSFF 275
 ++ K+ F +FG YRI+ G +LL+ FF
 Sbjct: 230 ISFLTNGFKVFGYRIIAGIILLIHHFIH 258

- 60 An alignment of the GAS and GBS proteins is shown below.

Identities = 227/272 (83%), Positives = 253/272 (92%)

Query: 1 MLIIEELLKALFLGVVEGVTEWLPVSSSTGHLILVQEFMKNQSKSFVEMFNIVIQLGAIMA 60
 MLIIEELLKA+F G++EG+TEWLPVSSSTGHLILVQEF++LNQ K+F+EMFNIVIQLGAI+A

-2107-

Sbjct: 1 MLIIELLKAIFFGIIEGITWLPVSSTGHLILVQEFIRLNQDKAFIEMFNIVIQLGAIIA 60

Query: 61 VIVIFYKRLNPFQPGKSAREIRLTWQLWLKVVIACIPSILIALPFDNWFEAHFNMPIA 120
V++IYF+RLNPFQPGK+ARE++LTWQLWLKVVIACIPSILIA+P DNWFEAHF FM+PIA

5 Sbjct: 61 VMLIYFERLNPFQPGKTAREVQLTWQLWLKVVIACIPSILIAVPLDNWFEAHFYFMVPIA 120

Query: 121 IALIFYGFVFIWVEKRNAHLKPQVTELASMSYKTAFLIGCFQVLSIVPGTSRSGATILGA 180
IALI YG FIW+EKRNA +P VTELA MSYKTAF IGCQVLSIVPGTSRSGATILGA

10 Sbjct: 121 IALIVYGIAFIWIIEKRNAQQEPVTELARMSYKTAFFIGCFQVLSIVPGTSRSGATILGA 180

Query: 181 IIIGTSRSVAADFTFFLAIPTMFGYSGLKAVKYFLDGNVLSLDQSLILLVASLTAFVVSL 240
II+GTSR+VAADFTFFLAIPTMFGYSGLKAVK+FLDG+ L Q LILLVASLTAFVVSL

Sbjct: 181 IILGTSRTVAADFTFFLAIPTMFGYSGLKAVKFELDGHHLDAQVLILLVASLTAFVVSL 240

15 Query: 241 YVIRFLTDYVKRHDFTIFGKYRIVLGSLLILY 272
IRFLTDYVK+HDFTIFGKYRIVLGSLL++Y

Sbjct: 241 LAIRFLTDYVKKHDFTIFGKYRIVLGSLLLIY 272

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
20 vaccines or diagnostics.

Example 1865

A DNA sequence (GBSx1972) was identified in *Sagalactiae* <SEQ ID 5801> which encodes the amino acid sequence <SEQ ID 5802>. Analysis of this protein sequence reveals the following:

Possible site: 42

25 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.65	Transmembrane	494 - 510 (488 - 519)
INTEGRAL	Likelihood = -8.01	Transmembrane	263 - 279 (256 - 288)
INTEGRAL	Likelihood = -5.95	Transmembrane	25 - 41 (20 - 43)
INTEGRAL	Likelihood = -4.94	Transmembrane	475 - 491 (473 - 493)

30 ----- Final Results -----

bacterial membrane	---	Certainty=0.4461(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

35

A related GBS nucleic acid sequence <SEQ ID 9481> which encodes amino acid sequence <SEQ ID 9482> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:AAB99606 GB:U67598 M. jannaschii predicted coding region MJ1577
[Methanococcus jannaschii]

Identities = 41/172 (23%), Positives = 78/172 (44%), Gaps = 19/172 (11%)

Query: 479 LISFVVIITYTLFLNYFTYFCIYLLLFQVILLNLIIFMMTRKISNGYIVTEDGASRVYQW 538
+IS ++ ++ F+ ++ + ++ ++ II +T G ++ +W

45 Sbjct: 442 VISILLAVFLYFIPKYSQTFNEVFYLSIVFVQNIILALTPTSLFGRWKANYKKEKL-EW 500

Query: 539 TSFRNMLRDIKSFDRSELESIVLWNRILVYATLFGYADRVEKALR-VNQIDIPERFANID 597
+F+N L ++ + E I +W L+Y T G D+V +A++ +N ++ + I

50 Sbjct: 501 DAFKNFLSNLAMIKKYSPEDISIKWDLIYGTALGVGDKVVEAMKSLNLSELVADYVVIH 560

Query: 598 SHQFAISVNQSSNHFTTITVDVSHASNFSVNSGGSSGGFSGGGG--GGGGGA 647
S+ ++ + S + ST GS GGF GGG GGGGGA

Sbjct: 561 SNYDSMKTSDSVYSSTT-----GSGGGFGAGGGFGGGGGGA 597

55 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5803> which encodes the amino acid sequence <SEQ ID 5804>. Analysis of this protein sequence reveals the following:

Possible site: 21
>>> Seems to have a cleavable N-term signal seq.

-2108-

INTEGRAL	Likelihood = -7.91	Transmembrane	486 - 502 (483 - 508)
INTEGRAL	Likelihood = -5.89	Transmembrane	465 - 481 (460 - 483)
INTEGRAL	Likelihood = -2.18	Transmembrane	244 - 260 (241 - 260)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

>GP:AAB99606 GB:U67598 M. jannaschii predicted coding region MJ1577
[Methanococcus jannaschii]
Identities = 59/263 (22%), Positives = 106/263 (39%), Gaps = 14/263 (5%)

15 Query: 369 FLDMAFGNKVTLFVDQLFSQYHYDADTIKQLKTKYKGGKLEQEVRSSEQVIKAMKKASA 428
++ + G K+ + L + Y++D +K L K K + E +S Q K+ K
Sbjct: 346 YIKIMNGGKIEILKTDLENLDVYESDVMKFLMKYSKNNVFDPEYIKSLAQKYKSSKDKLK 405

20 Query: 429 AITNNVLETIKKLNLPDITYRQMTPA--EKRSNSVQGLGCLLLILNSGLLIYLAIKESGL 486
+ + E K + P ++ A E R + L + ++L L ++
Sbjct: 406 KLKD---ELDKIMEYPRYSSKVVNAFLETRGKIIIIALLVISILAVFLYFIPKYSQTFN 462

25 Query: 487 ALIYLALMVLTMCLGFYISLKLQYKKGLETPEGGVRLHQWQSFKNMIRIDIDKFEDVAI 546
+ YL+++ + I L L G +W +FKN + ++ + +
Sbjct: 463 EVFYLSIVFVQ---NIIALTPTSLFGRWKANYKKELEWDAFKNFLSNLAMIKKYSP 518

30 Query: 547 EGLVWNRVLVYATLFGYAKKVERYLKVHRIALPEVYQAVRPGELSMVMYATPTTFVSSL 606
E + +W L+Y T G KV +K ++ + V + Y + T V S+
Sbjct: 519 EDISIWKDNLITGALGVGDKVVEAMKSLNLS-----ELVADYVIHSNYDSMKTSVDSV 573

35 Query: 607 SSATTSSNFSVSSGGGISGGGGG 629
S+TT S +GGG GGGGG
Sbjct: 574 YSSTTGSGGGFGAGGGFSGGGGGG 596

35 An alignment of the GAS and GBS proteins is shown below.

Identities = 241/635 (37%), Positives = 372/635 (57%), Gaps = 18/635 (2%)

40 Query: 22 MKKCFLAICLALSFVMSVQADEVDYNIPHYEGNLTIHNDNSADFTKVTYQFDSSYNGQ 81
MKK + + L S + ++A +VDY+I +YEG L + +N+A F +KVTYQFD+SYNGQ
Sbjct: 1 MKKILMTLVLCFSLGIRIKAADVDYSITNYEQQLLSKENTARFEQKVTYQFDTSYNGQ 60

45 Query: 82 YVTLTAGKLPDNFDINNKPQVEVSINGKVRKVSQIEDLEDGYRLKVFNGGEAGDTVKV 141
Y++LG G LP F I+ KP+VEV NG+ VS + DL DGYRLK++N G+AGD V V
Sbjct: 61 YISLGRGTGHLPAIDQKPKVEVYQNGQQVPVSQEFSDLGDRYRLKLYNAGQAGDKVDV 120

Query: 142 NVQWKLKNVLFMHKDVGELNWIPISDWDKTEKVDWFISTDKKVALSRLWGHLYL-KTP 200
V W+L ++L ++DV ELNW PISDWDKTEKV ++T + S LW H GY K P
Sbjct: 121 KVIWQLHLLLTAYQDVAELNWIPISDWDKTEKVSFTVTTPTDIQDSNLWAHRYGYQKQP 180

50 Query: 201 PKIRQNNRYHLTAFNVNKRLEFHGYWDRSYF--NLPTNSKNYKKKIEYQEKMIERHGF 258
+++ N+RY + A NV+ +LE H YWD+ P + + K KI E I R
Sbjct: 181 QVLKEGNSRYQINAKNVSGQLELHAYWDKKALLGKEPVDVSTSKKNKIVALETKISRRT 240

55 Query: 259 ILSFLLRILLPSFFIIVTLFISIRVFLFRKKVNKYGQFPKEHHLYEAPEDLSPIELTQSI 318
+L L ++P + L+ I+ +K+ N+Y H YE PEDLSPL LTQ+I
Sbjct: 241 LLQLLFGKVIPLVEVGFLWQLIQFTRLKQFNRYHLNHTDHSYEVPEPDLSPILVLTQAI 300

60 Query: 319 YSMSFKNFQ---DEEKKTHL---ISQEQLIQSILLDLIDRKVL---NYDDNLLSLANLD 368
Y SF E +K + ++ E L+Q+ LLDLID+KVL L ++ LD
Sbjct: 301 YGQSFAVLSPTASESQKLLIPKGVTFEALVQATLLDLIDQKVLTLTKKEGKAYLEISQLD 360

Query: 369 RASDAEIDFIEFADFSTSLKPDQLFSNYQFSYKETLRELKKQHKASDLQTMRRRGSNA 428
R +D E F++ AF + +L DQLFS Y + +T+++LKK +K L+ ++R+
Sbjct: 361 RVTDEEAFLDMAFGNKVTLFVDQLFSQYHYD-ADTIKQLKTKYKGGKLEQEVRSSEQV 419

65 Query: 429 LSRITRLTRLISKDNINSLRRKGISSPYRKMSSEESKELSRKRFSYLSPLISFVVIITYT 488

-2109-

```

      + + + + I+ + + + + + + YR+M+ E + + + + L + + + + +IY
Sbjct: 420 IKAMKASAAITNNVLETIKKLNLPDITYRQMTPEAKRKSNVQGLGCLLLILNSGLLIY- 478

Query: 489 LFLNYFTYFCIYLLLFVILLNKKIIFMMTRKISNGYIVTEDGASRVYQWTSFRNMLRDI 548
      L +          IYL L + + L I + +          I T +G R++QW SF+NM+RDI
Sbjct: 479 LAIKESGLALIYLLALMVLTMCLGFYISLKLQYKKLGIETPEGGVRLHQWQSFQNMIRDI 538

Query: 549 KSFDRSELESIVLWNRLVYATLFGYADRVKALRVNQIDIPERFANIDSHQFAISVNQS 608
      F+ +E +V+WNR+LVYATLFGYA +VE+ L+V++I +PE + + + + + +
Sbjct: 539 DKFEDVAIEGLVWNRVLVYATLFGYAKKVERYLKVHRIALPEVYQAVRFGELSMVMYAT 598

Query: 609 SNHFSSTITEDVSHASNFSVNSGGSSGGFSGGGGGG 643
      + F +          + +SNFSV+GG GG SGGGGGG
Sbjct: 599 TPTFVSSLSSATTSSNFSVSSG---GGISGGGGGG 630

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8921> and protein <SEQ ID 8922> were also identified. Analysis of this protein sequence reveals the following:

```

20 Lipop: Possible site: -1 Crend: 4
   McG: Discrim Score: 10.29
   GvH: Signal Score (-7.5): 3.11
      Possible site: 23
   >>> Seems to have a cleavable N-term signal seq.
25 ALOM program count: 3 value: -8.65 threshold: 0.0
      INTEGRAL Likelihood = -8.65 Transmembrane 475 - 491 ( 469 - 500)
      INTEGRAL Likelihood = -8.01 Transmembrane 244 - 260 ( 237 - 269)
      INTEGRAL Likelihood = -4.94 Transmembrane 456 - 472 ( 454 - 474)
      PERIPHERAL Likelihood = 2.28 540
30 modified ALOM score: 2.23

   *** Reasoning Step: 3

   ----- Final Results -----
35 bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no homology with any sequences in the databases.

40 Example 1866

A DNA sequence (GBSx1973) was identified in *S. agalactiae* <SEQ ID 5805> which encodes the amino acid sequence <SEQ ID 5806>. This protein is predicted to be glutamine-binding periplasmic protein/glutamine transport system perme. Analysis of this protein sequence reveals the following:

```

45 Possible site: 24
   >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL Likelihood = -8.86 Transmembrane 301 - 317 ( 295 - 324)
      INTEGRAL Likelihood = -6.05 Transmembrane 479 - 495 ( 473 - 496)
      INTEGRAL Likelihood = -0.59 Transmembrane 369 - 385 ( 369 - 385)

50 ----- Final Results -----
      bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

55 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAA17584 GB:D90907 glutamine-binding periplasmic protein
[Synechocystis sp.]
Identities = 147/534 (27%), Positives = 256/534 (47%), Gaps = 75/534 (14%)

```

-2110-

Query: 4 ILLSLFTALLITFGGMTSIQADEYLVRGMEAAAYAPFNWTQNDNINGAVPIEGTDQYANGY 63
 +LL++ LL F ++ + + V E + PF T E T Q G+
 Sbjct: 24 VLLAIAIPLLPAPFSQVSR---QTIIVATEPTFPFEMTD-----EATGQLT-GF 68

Query: 64 DVQVAKKLAKLKNKKVVVVKTKWEGLVLPALTSGKLDMIAGMSPTEERKKEINFSPYYI 123
 DV + + + + V + ++G++PAL S + I+ ++ T ER + ++FS PY+
 Sbjct: 69 DVDLIQAIGEAQVTVDIQGYPFDDGIIPALQSNVTGAASAITITPERAQSVSFSPYFK 128

Query: 124 SEPTLVVNABGKYTNAKNISDFKNAKVTAQQGVLYNLIDQINGVKKEVAMGDFNQLRQA 183
 S L + + KN+ D + ++ G + + G K + +F+ + A
 Sbjct: 129 S--VLAIAVQDGNDTIKNLKDLGKRLAVAIGTTGAMVATNVPGAK---VTNFDSTISA 182

Query: 184 VE---SGVVDAYVSERPDATSAQTANPKLKMIELHQGFKTSADDTNISVGMKRGDNRIHQ 240
 ++ +G DA ++RP A + L+ +++ + D I++ + INQ
 Sbjct: 183 LQELVNGNADAVINDRPVLLYA-IKDAGLRNVKISADVGSSEY-YGIAMPLAP-PGEINQ 239

Query: 241 VNQVL-----ESISRDKQIALMDKMIKEQ-----PSV-----KKEKNGK 274
 +VL + I A+ +K E+ PS+ + + N
 Sbjct: 240 TREVLNQGLFQIIENGTYNAIYEKWFGEKNPPFLPLVAPSLVGKVGTAQSLTERSQANPN 299

Query: 275 PNFFEQMATILKNNGSQFLRGTTATLLISMVGTIVGLFIGLLIGVFRTPAKSDNKLKAAL 334
 NF + T+ +N +G+ T+L++ GL G + + A SD
 Sbjct: 300 DNF---LITLFRN---LFGKSILTLLTAFSVFFGLIGGTGVAI---ALISD----- 342

Query: 335 QKLLGWLNNIYIEVFRGTPMIVQSMVIYYGTAQAF-----GVSLDRTLAAIFIVSINTGA 389
 K L + IY+E FRGTPM+VQ +IY+G F G+++DR AAI +S+N A
 Sbjct: 343 IKPLQLIFRIYVEFFRGTPMLVQLFIIYFGLPALFKEIGLITIDRFPAAIIALSLNVAA 402

Query: 390 YMSEIVRGGIFSVDKGQFEAATALGFTHCQTMRKIVLPQVVRNLPATGNEFVINIKDTS 449
 Y++EI+RGGI S+D+GQ+EA +LG + QTM++++ PQ R ILP GNEF+ IKDTS
 Sbjct: 403 YLAETIRGGIQSIDQGWEEACESLGMSFPWQTMKEVIFPQAFRRILPPLGNEFITLIKDTS 462

Query: 450 VLNVISVVELYFSGNTVATQTYQYFQFTTIIAIYFILFTVTRILRYIEKRPD 503
 + VI EL+ G + TY+ F+ + +A++Y +LT + + +++E D
 Sbjct: 463 LTAVIGFQELFREGLIVATTYRAFEVYIAVALVYLLTTISSFVFKWLENYMD 516

There is also homology to SEQ ID 1194.

A related GBS gene <SEQ ID 8923> and protein <SEQ ID 8924> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 2
 McG: Discrim Score: 6.23
 GvH: Signal Score (-7.5): 0.11
 Possible site: 24

>>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 3 value: -8.86 threshold: 0.0
 INTEGRAL Likelihood = -8.86 Transmembrane 301 - 317 (295 - 324)
 INTEGRAL Likelihood = -6.05 Transmembrane 479 - 495 (473 - 496)
 PERIPHERAL Likelihood = 1.32 441
 modified ALOM score: 2.27

*** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

34.3/57.3% over 462aa

EGAD|48193| glutamine-binding periplasmic protein/glutamine transport system permease
 protein Insert characterized

Synechocystis PCC6803

GP|1652664|dbj|BAA17584.1||D90907 glutamine-binding periplasmic protein {Synechocystis
sp.} Insert characterized
PIR|S77250|S77250 hypothetical protein - Synechocystis sp. (strain PCC 6803) Insert
characterized

EGAD|48193|sll1270(54 - 516 of 530) glutamine-binding periplasmic protein/glutamine transport system permease protein {Synechocystis PCC6803}GP|1652664|dbj|BAA17584.1||D90907 glutamine-binding periplasmic protein {Synechocystis sp.}PIR|S77250|S77250 hypothetical protein - Synechocystis sp. (strain PCC 6803)

Matches = 128 Mismatches = 149 Conservative Sub.s = 86

MKGMVKLGHWGKTWRYYL L LALGVLLAIAIPLLPAFSQVS
 10 20 30 40

444 474 495 525 555 585 615 645
FNWTFQNDNTNGAVPIEGTDC---YANGYDVQVAKKIAKKLNKKVVVKTKWEGLVPALTSGLDMIAGMSPTEERKKH
| | || :|| ::
RQTIIIVATEPTFFPFEMTDDEATGQLTGFDVDLIQAIGEAQVTVDIQGYFDFGIIPALQSNTVGAAISAITITPERAQSV
50 60 70 80 90 100 110 120

675 705 735 765 795 825 855 885
NFSKPYIISEPTLVVNABGKYNAKNISDFRNAKVTAQOGVLYNLIDQINGVKKEVAMGDFNQLRQAVSGGVDDAYVSE
: || | |: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
SFSSPYFKSVLAIAVQ-DGNDT-IKNLKDLEGRKLRAVAIGTTGAMVATNVPGAKVTNFDSITTSALQELV-NGNADAVIND

130 140 150 160 170 180 190

903 957 987
RP-----DATSAQTANPKLK-MIELHQG-FKTSADDTNISV
|| ||| :| :| :| :|
RPVLLYAIKDAGLRNVKISADV~~~NPPFLPLVAPSLVGKVGTAQSLTERSQANPNDNFILITLFRNLFKGS-----
210 270 280 290 300 310

```

1017      1047      1077      1107      1137      1167      1197      1227
GMRKGDNRINQVNOVLESISRDKQIALMDKMIKEQPSVKKEKNGKPNFFEQMATILKNNGSQFLRGATATLLISMVGTV
                                         : : : : |
-----ILTVLLTAF
                                         320

```

1257 1284 1314 1344 1374 1404 1419 1449
GLFIGLLIGV-FRTAPKSDNKLKAALQKLLGWLLNIYIEVFRGTPMIVQSMVIYYGTAQAF-----GVSLDRTLAAIFIV
:|:||:| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
SVFFGLIGGTGVAIALISD-----IKPLQLIFRIYVEFFRGTPMLVOLFIIFYGLPALFKEIGLGITIDRFPAAIAL

 340 350 360 370 380 390

[illegible]

1719 1749 1779 1809 1839 1869 1899 1929
 NTVATQTYQYQPTFTTIIAIYFILTFTVTRILRYIEKRFDSNYTTGANQLQV*EVGMTQAILIKHLKKSYSGSNEVLKT
 : | | : | : : : | | : : : : | :
 QLIVATTYRAFEVYIAVALVYLLTTTISFVFKWLENYMDPIGRAKKKAKAATA
 490 500 510 520 530

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2112-

Example 1867

A DNA sequence (GBSx1974) was identified in *Sagalactiae* <SEQ ID 5807> which encodes the amino acid sequence <SEQ ID 5808>. This protein is predicted to be ATP-binding. Analysis of this protein sequence reveals the following:

5 Possible site: 44
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.3208(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:CAB73160 GB:AL139076 putative glutamine transport ATP-binding
 protein [Campylobacter jejuni]
 Identities = 132/241 (54%), Positives = 178/241 (73%), Gaps = 1/241 (0%)

Query: 5 ILEIKHLKKSYSNEVLKDISLVNKGEVISIIGSSSGSKSTFLRSINLLEEPSGGEILY 64
 ++E+K+L+K YG EVLK+I+ +++KG+VI+IIG SG GKSTFLR IN LE GEIL
 20 Sbjct: 1 MIEVKNLQKKYGELEVLKNINTTISKGDVIAIIGPSGGKSTFLRCINRLELADSGEILI 60

Query: 65 HGHNVLKKGVDLNNYREKLGVMVFQSFNLFENLNILENAIVAQTTLKREKQEAKEIAKEN 124
 + N+L+K D+N R+K+ MVFQ FNLF N N++EN + ++EA K AK
 25 Sbjct: 61 NKQNILDKEIDINKIRQKVMVFQHFNLFANKNVNENLCLTPIKTGILSQEEAIKKAKLL 120

Query: 125 LNAVGMTEQYWKAKPKQLSGGQKQVVAIARALSVNPEAILFDEPTSAIDPEMVGEVLKTM 184
 L VG+ ++ P +LSGGQKQR+AIAR+L +NP+ ILFDEPTSAIDPEM+GEVL M
 30 Sbjct: 121 LAKVGLADKE-NIMPHKLSGGQKQRIAIARSLMMNPVILFDEPTSAIDPEMIGEVL SIM 179

Query: 185 QDLAKSGLTMIIVTHEMEFAKEVSDRVIFMDKGIIEAQCTPKQLFENPTQERTKEFLQRF 245
 +D+AK GLTM++VTHEM FA+ V++R+ FMDKG IA +PK++FENP+ ER +EFL + L
 35 Sbjct: 180 KDVAKGLTMLVVTHEMGFARNVANRIFMDKGLIADVASPKVEFENPNSNERLREFLNKVL 240

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2157> which encodes the amino acid
 35 sequence <SEQ ID 2158>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 40 bacterial cytoplasm --- Certainty=0.1170(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

45 Identities = 212/246 (86%), Positives = 237/246 (96%)

Query: 1 MTQAILEIKHLKKSYSNEVLKDISLVNKGEVISIIGSSSGSKSTFLRSINLLEEPSGG 60
 M+ +I+EI+L+KKSYSNEVLKDISLVNKGEVISIIGSSSGSKST LRSINLLEEPS G
 50 Sbjct: 24 MSNSIIEIKNLKKSYSNEVLKDISLVNKGEVISIIGSSSGSKSTLLRSINLLEEPSAG 83

Query: 61 EILYHGHNVLKKGVDLNNYREKLGVMVFQSFNLFENLNILENAIVAQTTLKREKQEAKEI 120
 +IL+HG +VL + Y+L +YREKLGVMVFQSFNLFENLN+LENAIVAQTTLK+R +AE+I
 55 Sbjct: 84 QILFHGEDVLAEHYNLTHYREKLGVMVFQSFNLFENLVLENAIVAQTTLKRDRAQAEQI 143

Query: 121 AKENLNAVGMTEQYWKAKPKQLSGGQKQVVAIARALSVNPEAILFDEPTSAIDPEMVGEV 180
 AKENLNAVGMTEQYW+AKPKQLSGGQKQVVAIARALSVNPEA+LFDEPTSAIDPEMVGEV
 60 Sbjct: 144 AKENLNAVGMTEQYWAQKPKQLSGGQKQVVAIARALSVNPEAMLFDEPTSAIDPEMVGEV 203

Query: 181 LKTMQDLAKSGLTMIIVTHEMEFAKEVSDRVIFMDKGIIEAQCTPKQLFENPTQERTKEF 240
 LKTMQDLAKSGLTMIIVTHEMEFA++VSDR+IFMDKG+I E+G+P+Q+FENPTQ+RTKEF
 60 Sbjct: 204 LKTMQDLAKSGLTMIIVTHEMEFARDVSDRIIFMDKGLITEGSPQIIFENPTQDRTKEF 263

-2113-

Query: 241 LQRFLK 246
 LQRFLK
 Sbjct: 264 LQRFLK 269

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1868

A DNA sequence (GBSx1976) was identified in *S.galactiae* <SEQ ID 5809> which encodes the amino acid sequence <SEQ ID 5810>. This protein is predicted to be hypersensitive-induced response protein. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -17.94 Transmembrane 4 - 20 (1 - 28)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.8175(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9479> which encodes amino acid sequence <SEQ ID 9480> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GF:AAF68390 GB:AF236374 hypersensitive-induced response protein
 [Zea mays]
 Identities = 127/275 (46%), Positives = 174/275 (63%), Gaps = 1/275 (0%)
 Query: 19 ITSLYVVKQQTVAIIERFGKYQKTATSGIHIRVPLGIDKIAARVQLRLLQSEIIVETKTK 78
 I L V Q T V A I E F G K + + G H + I A + L R + Q + + E T K T K
 Sbjct: 4 ILGLVQVDQSTVAIKENFGKFSEVLEPGCHFLPWCIGQQIAGYLSLRVRQLDVRCECTKTK 63
 Query: 79 DNVFVTLNIATQYRVNENNVTDAYYKLIKPEAQIKSYIEDALRSSVPKLTLDLFEKKDE 138
 DNVFVT+ + QYR + +DA+YKL QI+SY+ D +R++VPKL LD+ FE+K+E
 Sbjct: 64 DNVFVTVVASVQYRALADKASDAFYKLSNTRREQIQSYVFDVIRATVPKLGLDLDAFEQKNE 123
 Query: 139 IALEVQHQAEEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQRKRVAQAELANADKI 198
 I A V + + + M S T Y G Y I V + T L I + E P D V K + + M N E I N A A R V A A E A A + K I
 Sbjct: 124 IAKAVEEELEKAMSTYGYQIVQTLIVDIEPDDRVKRAMNEINAAARMRVAASEKAEAEKI 183
 Query: 199 KIVTAAEAABAEKDRLHGVGIAQQRKAIVDGLADSIQELKDANVTLTTEEQIMSILLTNQYL 258
 + A E E A E L G V G I A + Q R + A I V D G L D S + + T + I M + + L Q Y
 Sbjct: 184 LQIKKAEGEAESKYLAVGVIARQQAIVDGLRDSVLAFSENVPGTTAKDIMDMVLVTQYF 243
 Query: 259 DTLNIF-AINGNQTIFLPNNPEGVEDIRTQVLSAL 292
 D T + A + + + F + P + P V + D + Q + L
 Sbjct: 244 DTMREIGASSKSSSVFIPHGPAVKDVSQAIRDGL 278

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5811> which encodes the amino acid sequence <SEQ ID 5812>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -13.06 Transmembrane 5 - 21 (1 - 29)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.6222(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-2114-

The protein has homology with the following sequences in the databases:

```
>GP:AAF68390 GB:AF236374 hypersensitive-induced response protein
[Zea mays]
Identities = 126/273 (46%), Positives = 174/273 (63%), Gaps = 3/273 (1%)

5   Query: 23  LYVVRQQSVAIVERFGRYQKTATSGIHIRLPFGI-DKIAARVQLRLQSEIIVETKTKDN 81
      L V Q +VAI E FG++ +   G H LP+ I +IA + LR+ Q ++ ETKTKDN
      Sbjct: 7   LVQVDQSTVAIKENFGKFSEVLEPGCHF-LPWCIGQQIAGYLSLRVRQLDVRCECTKTKDN 65

10  Query: 82  VFVTLNVAIQYRVNEQNVTDAYYKLMKPESQIKSYIEDALRSSVPKLTDELFEKKDEIA 141
      VFVT+  + QYR      +DA+YKL      QI+SY+ D +R++VPKL LD+ FE+K+EIA
      Sbjct: 66 VFVTVVASVQYRALADKASDAFYKLSNTRERQIQSYVFDVIRATVPKLGLLDAFEQKNEIA 125

15  Query: 142 LEVQHQAEEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQRKRVAAQELANADKIKI 201
      V+ ++ + MSTYGY IV+TLI +EPD VK++MNEINAA R RVAA E A A+KI
      Sbjct: 126 KAVEBELEKAMSTYGYQIVQTLIVDIEPDDRVRKRAMNEINAAARMVAASEKAEAEKILQ 185

20  Query: 202 VTAAEAAEAKDRHLHGVGIAQQRKAIVDGLAESIQELKEANISLNEEQIMSILLTNQYLD 261
      + AE EAE L GVGIA+QR+AIVDGL +S+ E      + IM ++L QY DT
      Sbjct: 186 IKKAEGEAESKYLAVGIAQRQAIVDGLRDSVLAFSENVPGTTAKDIMDMVLVTQYFDT 245

      Query: 262 LNTFAAKG-NQTLFLNTPSGVEDIRTOVLSAL 293
      + A + ++F+P+ P V+D+ Q+ L
      Sbjct: 246 MREIGASSKSSSVFIPHGPAGVAKDVSAQIRDGL 278
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 254/291 (87%), Positives = 278/291 (95%)

30  Query: 5   IILTIVLVIVLLITSLYVVKQQTVAIIERFGKYQKTATSGIHIRVPLGIDKIAARVQL 64
      I +   +++++ ++ ++LYVV+QQ+VAI+ERFG+YQKTATSGIHIR+P GIDKIAARVQL
      Sbjct: 6   IFIAPGVIVILAIVASTLYVVRQQSVAIVERFGRYQKTATSGIHIRLPFGIDKIAARVQL 65

      Query: 65  RLLQSEIIVETKTKDNVFTLNIAIQYRVNENNVTDAYYKLIKPEAQIKSYIEDALRSSV 124
      RLLQSEIIVETKTKDNVFTLN+ATQYRVNE NVTDAYYKL+KPE+QIKSYIEDALRSSV
35  Sbjct: 66  RLLQSEIIVETKTKDNVFTLNVAIQYRVNEQNVTDAYYKLMKPESQIKSYIEDALRSSV 125

      Query: 125 PKLTDELFEKKDEIALEVQHQAEEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQR 184
      PKLTDELFEKKDEIALEVQHQAEEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQR
40  Sbjct: 126 PKLTDELFEKKDEIALEVQHQAEEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQR 185

      Query: 185 KRVAQAELANADKIKIVTAAEAAEAKDRHLHGVGIAQQRKAIVDGLADSIQELKANDVTLT 244
      KRVAQAELANADKIKIVTAAEAAEAKDRHLHGVGIAQQRKAIVDGLA+SIQELK+AN++L
      Sbjct: 186 KRVAQAELANADKIKIVTAAEAAEAKDRHLHGVGIAQQRKAIVDGLAESIQELKEANISLN 245

45  Query: 245 EEQIMSILLTNQYLDTLNTFAINGNQTFIFLNNPEGVEDIRTOVLSALKTR 295
      EEQIMSILLTNQYLDTLNTFA GNQT+FLFN P GVEDIRTOVLSALKT+
      Sbjct: 246 EEQIMSILLTNQYLDTLNTFAAKGNQTLFLNTPSGVEDIRTOVLSALKTK 296
```

SEQ ID 5810 (GBS231) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 7; MW 60.9kDa).

GBS231d was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 5-7; MW 59kDa) and in Figure 239 (lane 11; MW 59kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 9; MW 34kDa) and in Figure 183 (lane 6; MW 34kDa). Purified GBS231d-GST is shown in Figure 246, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2115-

Example 1869

A DNA sequence (GBSx1977) was identified in *S.agalactiae* <SEQ ID 5813> which encodes the amino acid sequence <SEQ ID 5814>. Analysis of this protein sequence reveals the following:

```

Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2305(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9291> which encodes amino acid sequence <SEQ ID 9292> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:CAB13457 GB:Z99112 similar to hypothetical proteins [Bacillus subtilis]
    Identities = 259/514 (50%), Positives = 350/514 (67%), Gaps = 9/514 (1%)

Query: 1  MGMTMENGAKEVSDKPATTVGEVQILSKGVLGMGARGNSGVITSQLFRGFGQSIKDKBEL 60
      M ++M +GA+EV +G+VG LSKG+LMGARGNSGVI SQLFRGF ++I+ K+E+
20  Sbjct: 46  MNLSMTSGAREVEQMDTDDIGKVGSAKGLLMGARGNSGVILSQLFRGFSKNIETKKEI 105

Query: 61  TGQDLAHAFQNGVEVAYKAVMKPVEGTILTVSRGAATAALKKAEETDDAVEVMRATLKGA 120
      + A A Q GV++AYKAVMKPVEGTILTV++ AA A+ AE+ D +M A + A
25  Sbjct: 106 NALEFAAALQAGVDMAYKAVMKPVEGTILTVAKDAKKAMILAEKETDITALMTAVTEEA 165

Query: 121 KRALAKTPDMLPVLKEVGVDGSGGGLVFIYEGFLSALTGEYIASDFKATPATMTEMVN 180
      + +L +TP++LPVLKEVGVDGSGG+GL+ +YEGFL++L GE + KA ++ +MV+
30  Sbjct: 166 EASLNRTPELLPVLKEVGVDGSGGKLLCVYEGFLASLKGETVPQ---KAVLPSLDDMVS 222

Query: 181 AEHHKAVVGHVATEDIKGYCTEVMVGLKQGPTVYKFEFNYEEFQGYLSNLGDSLLVVDND 240
      AEHHK+ + TEDI++G+CTEVMV L Q +EF+ F+ LS GDSLLV+ D+
35  Sbjct: 223 AEHHKSAQSMNTEDIEFGFCTEVMVRLDQTK---REFDEGTFRQDLSQFGDSSLVIAD 279

Query: 241 EIVKVHVHTEDPGLVMQEGCLKYKSLVKVVENMRNQHDA---QMCKVEVEETVKETKEYG 297
      + KVV+H E+PG V+ YG L+K+K+ENMR QH + Q K ET + YG
40  Sbjct: 280 SLAKVHIHAEEPGNVNLNYAQHYGELIKIKIENMREQHTSIISQESKPADNETPPAKQPYG 339

Query: 298 IIAVVAGDGLAEIFKSQGVDIYIISGGQTMNPSTEDIVKAIEKVNARNVILPNNKNIFMA 357
      I+ V G+G+A++FKS G +I GGQTMNPSTEDIV A++ VNA V ILPNN NI MA
45  Sbjct: 340 IIVTAMGEGIALDFKSGASVVEGGQTMNPSTEDIVDAVKSVDNADTVFILPNNNSNIIMA 399

Query: 358 AQSAAADVVDIPAAVVETRTVPQGFTSLAFLDPAKSLETNVADMTNSLSDVISGSVTLAVR 417
      A AA VVD V+ +TVPQG ++LLAF+P + E N A+M +++ V SG VT +VR
50  Sbjct: 400 ANQAASVVDEQVFVIPAKTVPQGMASALLAFNPQEAEEANEANMLSAIQQVKSQVTFVSR 459

Query: 418 DTTIDGLEIHENDILGMVDGKILVSTPDMEKALKDTFDMIDEDESEIVTIYVGEDGKQAL 477
      DT IDG +I + D +G+++G I+ ++ + A K +MI ED EIVTI GED Q
55  Sbjct: 460 DTHIDGDIKKGDFMGILNGTIIGTSENQLSAAKMLLSEMIGEDDEIVTILYGEDASQEE 519

Query: 478 AETLSEYLEETYEDVEVEIHQGDQPVYPYLSVVE 511
      AE L +L E YE++EVEIH G QP+Y Y++S E
60  Sbjct: 520 AEQLEAFLSEKYEEIEVEIHNGKQPLYSYIVSAE 553

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5635> which encodes the amino acid sequence <SEQ ID 5636>. Analysis of this protein sequence reveals the following:

```

Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1816(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

```

-2116-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 434/511 (84%), Positives = 475/511 (92%)

5 Query: 1 MGMTMENGAKEVSDKPATTVGEVQILSKGVLMGARGNSGVITSQLFRGFGQSIKDKEEL 60
 M MTM+NGAKEV+DKPA+TVGEVGG+LSKG+LMGARGNSGVITSQLFRGFGQSIK K+EL
 Sbjct: 44 MSMTMDNGAKEVADKPASTVGEVQMLSKGLLMGARGNSGVITSQLFRGFGQSIKGDDEL 103

10 Query: 61 TQDLAHAFQNGVEVAYKAVMKPVEGTILTVSRGAATAALKKAEETDDAVEVMRATLKGA 120
 TG+DLA AFQ GVEVAYKAVMKPVEGTILTVSRGAATAALKKA+ TDDAVEVM+A L GA
 Sbjct: 104 TGKDLAQAFQVGVEVAYKAVMKPVEGTILTVSRGAATAALKKADLTDDAVEVMAALDGA 163

15 Query: 121 KRALAKTPDMLPVLKEVGVDSCGGQLVFIYEGFLSALTGEYIASEDFKATPATMTEMVN 180
 K ALAKTPD+LPVLKEVGVDSCGGQLVFIYEGFLSAL G+Y+ S DFKATPA M+EM+N
 Sbjct: 164 KGALAKTPDLLPVLKEVGVDSCGGQLVFIYEGFLSALNGDYVTSADFKATPANMSEMIN 223

20 Query: 181 AEHHKAVVGHVATEDIKYGYCTEVMVGLKQGPTYVKEFNYYEEFQGYLSNLGDSLLVVND 240
 AEHHK+VVGHVATEDI YGYCTE+MV LKQGPTYVKEFNYY+EFQGYLS LGDSLLVVND
 Sbjct: 224 AEHHKSVVGHVATEDITYGYCTEIMVALKQGPTYVKEFNDEFQGYLSGLGDSLLVVND 283

25 Query: 241 EIVKVHVHTEDPGLVMQEGGLKYGSLVKVKNMRNQHDAQMQKVEVEETVKETKEYGIIA 300
 EIVKVHVHTEDPGLVMQEGGLKYGSL+K+KV+NMNRNQH+AQ+QK +VE+ E K++G+IA
 Sbjct: 284 EIVKVHVHTEDPGLVMQEGGLKYGLIKIKVDNMRNQHEAQVQKTDVEKNKAEVKDFGLIA 343

30 Query: 301 VVAGDGLAEIFKSQGVYIISGGQTMNPSTEDIVKAIEKVNARNVILPNNKNIFMAAQS 360
 VVAG+GL+EIFK+QGVYI+ISGGQTMNPSTEDIVKAIE VNA+ VIILPNNKNIFMAAQS
 Sbjct: 344 VVAGEGLSEIFKAQGVYVISGGQTMNPSTEDIVKAIEAVNAKQVILPNNKNIFMAAQS 403

35 Query: 361 AADVVDIPAAVVEVTRTPVQGFSTLLAFDPAKSLETNVADMTNSLSDVISGSVTLAVRDTT 420
 AA+VVDIPAAVV TRTPVQGFSTLLAFDP+KSLE NVADM+ SLSDV+SGSVTLAVRDTT
 Sbjct: 404 AAEVVDIPAAVVATRTVPCGFSTLLAFDPSKSLEDNVADMSTSLSDVVGSVTLAVRDTT 463

40 Query: 421 IDGLEIHENDILGMVDGKILVSTPDMEKALKDTFDMIDESEIVTIYVGEDGKQALAE 480
 IDGLEIHEND LGMVDGKI+VS PDME LK F+KMIDEDSEIVTI+VGE+G Q LAE
 Sbjct: 464 IDGLEIHENDFLGMVDGKIIVSNPDMEATLKAAFEKMIDEDSEIVTIFVGEEDQDLAE 523

45 Query: 481 LSEYLEETVEDVEVEIHQGDQPVPYPLMSVE 511
 L+ YL ETYEDVEVEIHQGDQPVPYPLMSVE
 Sbjct: 524 LAGYLGETYEDVEVEIHQGDQPVPYPLMSVE 554

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1870

45 A DNA sequence (GBSx1978) was identified in *S.galactiae* <SEQ ID 5815> which encodes the amino acid sequence <SEQ ID 5816>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4771(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2117-

Example 1871

A DNA sequence (GBSx1979) was identified in *S.agalactiae* <SEQ ID 5817> which encodes the amino acid sequence <SEQ ID 5818>. This protein is predicted to be proliferating-cell nucleolar antigen P120. Analysis of this protein sequence reveals the following:

5 Possible site: 55
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3774(Affirmative) < succ>
 10 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9345> which encodes amino acid sequence <SEQ ID 9346> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC74905 GB:AE000278 putative nucleolar proteins [Escherichia coli K12]
 Identities = 87/229 (37%), Positives = 128/229 (54%), Gaps = 8/229 (3%)

20 Query: 63 GKSIEHTTGLVYSQEPAAQ--IVAQIAEPQEGMKVLDLAAAPGGKTHLLSYLNNIGLLV 120
 G + EH +GL Y QE ++ + A A+ +V+D+AAAPG KTT + + +NN G ++
 Sbjct: 89 GSTAHLGLFYIQEASSMLPVAALFADGNAPQVRMDVAAAPGSKTTQISARMNNEGAIL 148

25 Query: 121 SNEISNKRKSKILVENVERFGARNVIVTNESSQRLAKCFNSFFDLIVFDGPGCSGEGMFRKD 180
 +NE S R K+L N+ R G NV +T+ + FD I+ D PCSGEG+ RKD
 Sbjct: 149 ANEFSASRVKVLHANISRCGISNVALTHFDGRVFGAAVPEMFDAILLDAPCSGEGVVRKD 208

30 Query: 181 PQAIQYWHKDYPTCAQLQRDILKEAIKMLAHGGILVYSTCTWSPEENEEVVNWLQEY- 239
 P A++ W + E A QR+++ A L GG LVYSTCT + EENE V WL + Y
 Sbjct: 209 PDALKNWSPEASNQETIAATQRELIDSAFHALRPGGTLVYSTCTILNQEENEAVCLWLKETYP 268

35 Query: 240 ---DYLELVDIPKLGMEVGINVPQVARMYPHHFQEGGFVAKLRDTRS 285
 ++L L D+ G + + ++P + EG FVA+LR T++
 Sbjct: 269 DAVEFLPLGDL--FPGANKALTEEGFLHVFPQIYDCEGFFVARLRKTQA 315

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5819> which encodes the amino acid sequence <SEQ ID 5820>. Analysis of this protein sequence reveals the following:

40 Possible site: 34
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2316(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 An alignment of the GAS and GBS proteins is shown below.

Identities = 213/311 (68%), Positives = 254/311 (81%), Gaps = 3/311 (0%)

50 Query: 1 MKLPNEFIEKYQTILKDEAEAFFDSFEQKPI SAYRTNPLKEKQLDFPNAIPSTPWGHY GK 60
 M LP EFI YQ IL E E F SF Q+P++A+R NPLK + F + IP+T WG+Y GK
 Sbjct: 2 MSLPKFEFINTYQAILGKELEDFLASFNQEPVNAFRINPLKNQLKTFEHPENTLWGY Y GK 61

55 Query: 61 ISGKSIEHTTGLVYSQEPAAQIVAQIAEPQEGMKVLDLAAAPGGKTHLLSYLNNIGLLV 120
 +SGKS EH +GLVYSQEPAAQ+VAQ+A PQ+G +VLDLAAAPGGK+THLL+YL+NTGLLV
 Sbjct: 62 LSGKSPHEVSGLVYSQEPAAQMVAAAPQKGSVLDLAAAPGGKSTHLLAYLDNTGLLV 121

 Query: 121 SNEISNKRKSKILVENVERFGARNVIVTNESSQRLAKCFNSFFDLIVFDGPGCSGEGMFRKD 180
 SNEIS KRSK+LVEN+ERFGARNV+VTNES+ RLAK F+ +FD IVFDGPGCSGEGMFRKD
 Sbjct: 122 SNEISKRSKVLVENIERFGARNVVVTNESADRLAKVFSHYFDTIVFDGPGCSGEGMFRKD 181

-2118-

Query: 181 PQAIQYWHKDYPTCAQLQORDILKEAIKMLAHGGILVYSTCTWSPEENEEVNVWLLQEYD 240
 P AIQYWH YP ECA+LQ+ IL++A+ ML GG L+YSTCTW+PEENE+VV WLL+ Y
 Sbjct: 182 PDAIQYWHHGYPACAKLQKSILEDALAMLKPGGELIYSTCTWAPENEDVVQWLLLETYT 241

Query: 241 YLELVDIPKLNMGVGINVPQVARMYPHFFQGGQFVAKLRDTRSKEAQKIKPKAQKIN- 299
 +LELVD+PKLNMGV GI +P+ ARMYPH +QGGQFVAKL+D R +E Q K KA K N
 Sbjct: 242 FLELVDVPKLNMGVSGIGLPETARMYPHRYQGGQFVAKLKDKR-QEGQSTKLKAPKSNL 300

Query: 300 -KMQLQLWQQF 309
 K QL+LW+ F
 Sbjct: 301 IKDQLRLWKMF 311

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1872

A DNA sequence (GBSx1980) was identified in *S.agalactiae* <SEQ ID 5821> which encodes the amino acid sequence <SEQ ID 5822>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4111(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC24940 GB:AF012285 unknown [Bacillus subtilis]
 Identities = 86/240 (35%), Positives = 133/240 (54%), Gaps = 10/240 (4%)

Query: 6 DFAKQLVYKAGQFIKSEMONTFDVEEKSFRDDLVTSLDKKTQKLLIQEIIQHYPDDNILA 65
 + AK+ + +AG I M + +E KS +DLVT++DK+T+K I I + +P IL
 Sbjct: 9 EIAKWIREAGARITQSMHESATITETKSNPNDLVNIDKETEKFFIDRIQETFFPGRHILG 68

Query: 66 EE---DBVRSPIAQGNVWVLDPIDGTNVFIVQKDNFAVMLAYYEEGVQFGIIYDVMADI 122
 EE D + S +G VW++DPIDGT+NF+ Q+ NFA+ + +E G G+ G+IYDV+ D
 Sbjct: 69 EEEGQDKIHS--LEGVVWIIDPIDGTMNFBVHQQRNFAISIGIFENGEGKIGLIYDVVHDE 126

Query: 123 LYSGGGHFDVYANDKKIVPFQECPLERCLLGVNSAMYAEN----DCGIAHLASETLGVRI 178
 LY Y N+ K+ P +E +E +L +N+ EN +A L G R
 Sbjct: 127 LYHAFSGRGAYMNETKLAPLKETVIEEAILAINATWVTENRRIDQSVLAPLVKRVGRTRS 186

Query: 179 YGGAGISMAKVMQKLLAYFSY-IQPWDYAAKIMGETLGFLLTLDGEEPNYSTRQKVM 237
 YG A + +A V G++ AY + + PWDYAA ++ +G T T++GE + V+
 Sbjct: 187 YGSAALELANVAAGRIDAYITMRLAPWDYAAGCVLLNEVGGTYYTTIEGEPFTFLENHSLV 246

A related GBS nucleic acid sequence <SEQ ID 10937> which encodes amino acid sequence <SEQ ID 10938> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5823> which encodes the amino acid sequence <SEQ ID 5824>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1843(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2119-

An alignment of the GAS and GBS proteins is shown below.

Identities = 155/253 (61%), Positives = 205/253 (80%)

```

5  Query: 1  MDAKFDFAKQLVYKAGQFIKSEMQNTFDVEEKSRFDDLVTSLDKKTQKLLIQETIQHYPD 60
    ++ K+ FA+Q++ +AG FIKS+M      D++ K++FDDLVT++D++TQ+LL+  I Q YP
    Sbjct: 8  LETKYAFARQIIKEAGLFIKSKMSEQLDIQVKTQFDDLVTNVDQETQQLMDRIHQTYPC 67

10 Query: 61  DNILAEEDBVRSPIAQGNVWVLDPIDGTVNFIVQKDNFAVMLAYYEEGVGQFGIYDVMA 120
    D ILAEE++VR PI QGNVWV+DPIDGTVNFIVQ  FAVM+AYYE+G+GQFG+IYDVMA
    Sbjct: 68  DAILAEENDVRHPINQGNVWVIDPIDGTVNFIVQGSQFAVMIAYYEQGIGQFGIYDVMA 127

15 Query: 121  DILYSGGGHFDVYANDKKIVPFQECPLERCLLGVNSAMYAENDCGIAHLASETLGVRIYG 180
    D L +GGG F+V N K+ +QE PLER L+G N+ M+A ND  +AHL ++TLGVR+YG
    Sbjct: 128  DQLLAGGGDFEVTNLNGDKLPAYQEKPLERSLIGCNAGMFARNDRLAHLIAKTLGVRVYG 187

20 Query: 181  GAGISMAKVMQKLLAYFSYIQPWDYAAAKIMGETLGTLLTLTGDEEPNYSTRQKVMFLP 240
    GAGI M KVM+ +LLAYFS+IQPWDYAAAK++G+ LG+ LLT+DG EP++ TRQK+MF+P
    Sbjct: 188  GAGICMVKVMQKELLAYFSFIQPWDYAAAKVLGDKLGYVLLTIDGYEPDFQTRQKIMFVP 247

    Query: 241  KSKLNLIQSYLTK 253
    K +L I S+LTK
    Sbjct: 248  KCQLTRIASFLTK 260
  
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1873

A DNA sequence (GBSx1981) was identified in *Sagalactiae* <SEQ ID 5825> which encodes the amino acid sequence <SEQ ID 5826>. Analysis of this protein sequence reveals the following:

```

30 Possible site: 16
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.4131(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
  
```

The protein has homology with the following sequences in the GENPEPT database.

```

40 >GP:AAC24938.GB:AF012285 unknown [Bacillus subtilis]
    Identities = 33/78 (42%), Positives = 50/78 (63%)

    Query: 13  YSYPLDPSWNTEDITKVLRFNLQVEHAYENSIKVDLLDSYKEFKKVVKSKAQEKQIDRE 72
    Y YP++ W TE+ V+ F QVE AYE      ++LL +Y+ FK++V KA+EK++ E
    Sbjct: 3  YQYPMNEDWTTEEAVDVIAFFQQVELAYEKGADRELLKAYRRFKEIVPGKAEEKKLCGE 62

45    Query: 73  FQRTSGYSTYQAVKAAQQ 90
    F+ S YS Y+ VK A++
    Sbjct: 63  FEEQSTYSPYRTVKQARE 80
  
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5827> which encodes the amino acid sequence <SEQ ID 5828>. Analysis of this protein sequence reveals the following:

```

55 Possible site: 60
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.4442(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
  
```

An alignment of the GAS and GBS proteins is shown below.

-2120-

Identities = 59/91 (64%), Positives = 70/91 (76%)

Query: 9 ISSNYSYPLDPSWNTEDITKVLRLNQEVEHAYENSIKVDLLDSYKEFKKVVVSKAQEKQ 68
 +S NY YPLD SW+TE+I+ VL FLN+VE AYE + LLDSYK +K +VKSKAQEKQ

5 Sbjct: 5 MSGNYYPLDLWSSTEEISSVLHFLNKVELAYEKKVDAKQLLDSYKTYKTIVKSKAQEKQ 64

Query: 69 IDREFORTSGYSTYQAVKAAQQQAKGFISLG 99

IDR+FQ+ SGYSTYQ VK A+ KGF SLG

10 Sbjct: 65 IDRDFQKVSGYSTYQVVKKAKAIEKGFFSLG 95

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1874

15 A DNA sequence (GBSx1982) was identified in *S.agalactiae* <SEQ ID 5829> which encodes the amino acid sequence <SEQ ID 5830>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence (or aa 1-18)

20 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.0952(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:AAF21893 GB:AF103794 unknown [Listeria monocytogenes]

Identities = 74/126 (58%), Positives = 101/126 (79%)

Query: 1 MITLFLSPSCTSCRKARAWLSKHEVAFEEHNIITSPLNKEELLQILSFTENGTEIISTR 60

30 M+TL+ SPSCCTSCRK+RAWL +H++ ++E NI + PL+ +E+ +IL TE+GT++IISTR

Sbjct: 1 MVTLYTSPSCTSCRKSRWLEEHDIPIKERNIFSEPLSLDEIKEILRMTEGTEIISTR 60

Query: 61 SKVFQKLAIIDVDELSTSSLMELISENPILLRRPIILDKKRMQIGFNEDEIRAFPRDYRK 120

SK FQKL +D+D L L ELI +NP LLRRPII+D+KR+Q+G+NEDEIR FLPR R

35 Sbjct: 61 SKTFQKLINVDLDSLPLQQLFELIQKNPGLLRPIIIDEKRLQVGYNEDEIRFLPRRVRT 120

Query: 121 QELKQA 126

+L++A

Sbjct: 121 YQLREA 126

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5831> which encodes the amino acid sequence <SEQ ID 5832>. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.0511(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 An alignment of the GAS and GBS proteins is shown below.

Identities = 112/134 (83%), Positives = 127/134 (94%)

Query: 1 MITLFLSPSCTSCRKARAWLSKHEVAFEEHNIITSPLNKEELLQILSFTENGTEIISTR 60

55 M+TLFLSPSCTSCRKARAWL KHEV F+EHNIIITSPL+++EL+ ILSFTENGTEIISTR

Sbjct: 1 MVTFLFLSPSCTSCRKARAWLVKHEVDFQEHNIITSPLSRDELSILSFTENGTEIISTR 60

Query: 61 SKVFQKLAIIDVDELSTSSLMELISENPILLRRPIILDKKRMQIGFNEDEIRAFPRDYRK 120

SKVFQKL IDV+ELS S L++LI++NPSILLRRPII+D+KRMQIGFNEDEIRAF R DYRK

Sbjct: 61 SKVFQKLDIDVEELSISDLIDLIAKNPSILLRRPIIMDQKRMQIGFNEDEIRAFSRDYRK 120

-2121-

Query: 121 QELKQATIRAEIEG 134
 QEL+QATI+AEIEG
 Sbjct: 121 QELRQATIKAEIEG 134

5

SEQ ID 5830 (GBS232) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 10; MW 16.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 2; MW 42kDa).

GBS232-GST was purified as shown in Figure 207, lane 7.

10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1875

A DNA sequence (GBSx1983) was identified in *S.galactiae* <SEQ ID 5833> which encodes the amino acid sequence <SEQ ID 5834>. Analysis of this protein sequence reveals the following:

15 Possible site: 39
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
 20 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5835> which encodes the amino acid sequence <SEQ ID 5836>. Analysis of this protein sequence reveals the following:

25 Possible site: 39
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.1768(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 210/308 (68%), Positives = 252/308 (81%)

35 Query: 1 MKIHYNIDYKDIQAKEDCVLVLGYFDGLHLGHKALFDKAKKIATEKNLKIIVLTFNETPR 60
 M+I YI DY+DI ++D VL+LG YFDGLH GHKALFDK A++A ++ LK+VV TF E+P+
 Sbjct: 1 MEIEYIKDYRDINQEDDTVLILGYFDGLHRGHKALFDKAREVANKEGLKVVVFTFTESPK 60

40 Query: 61 LTFARFQPELLLHLTSPEKRSEKQFYGVDEL YLMNFTSHFSKVSSDLFIKKYIYGLRAK 120
 L F+RF PELL LH+T P+KR EKF +YGV++LYL++FTS FSKVSSD FI YI L+AK
 Sbjct: 61 LAFSRFSPELLLHITYPKRYEKFADYGVNKL YLVDFTSKFSKVSSDHFITHYIKNLKAK 120

45 Query: 121 AAVVGFDYKFGHNRTSGDYLRNFKGFPVYIIDEISEGGEKISSTRIRQLITEGNVEKANQ 180
 VVGFDYKFGHNRT DYLRNF+G VY I+EI E KIS+T IR+LI EGNV KAN
 Sbjct: 121 HIVVGFDYKFGHNRTSDYLTRNFEGQVYTIIEIKEDHRKISATWIRKLIQEGNVVKANH 180

50 Query: 181 LLGYEFSTCGMVVHGDARGRTIGFPTANLAPINRTYLPADGVYISNVLINGKYRAMTSI 240
 LLGY+ ST G VVHGDARGRTIGFPTANLAPI+ TYLPADGVY++NV++ K YR+MTS+
 Sbjct: 181 LLGYDLSTRGRVVHGDARGRTIGFPTANLAPIDNTYLPADGVYVTVNIVANKIYRSMTSL 240

Query: 241 GKNITFGGTELRLLEANIFDFDGIYGETIEIFWLKRIEMVKFNGIDDLVKQLKKDKKEIA 300
 GKN+TFGG ELRL E NIFDFD +IYGE IEI WL +IR+M KF GI+DL +L+ DK A
 Sbjct: 241 GKNVTFGGKELRLVNI FDFDEEYIYGEIIEIVWLDKIRDMKFE GIEDLTDRL EYDKRTA 300

55 Query: 301 LNWKDSQ 308

-2122-

LNWKKDS+
Sbjct: 301 LNWKKDSK 308

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1876

A DNA sequence (GBSx1984) was identified in *S.agalactiae* <SEQ ID 5837> which encodes the amino acid sequence <SEQ ID 5838>. This protein is predicted to be tRNA pseudouridine 5S synthase (truB). Analysis of this protein sequence reveals the following:

Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2576(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9817> which encodes amino acid sequence <SEQ ID 9818> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06129 GB:AP001515 tRNA pseudouridine 5S synthase [Bacillus halodurans]
Identities = 145/283 (51%), Positives = 191/283 (67%), Gaps = 12/283 (4%)

Query: 2 ITGIINLKKKAGMTSHDAVFKLRLKILHTKKIGHGGTLDPDVVGVLPIAVGKATRVIEYMT 61
+TGI+ L K GMTSHD V KLR++L TKK+GH GTLDPDV GVLP+ +G AT+V +YM+
Sbjct: 3 MTGILPLAKPRGMTSHDCVAKLRRLLKTKKVGHTGTLDPDVVGVLPCIGHATKVAQYMS 62

Query: 62 ESGKIYEGEITLGYATSTEDSSGEVISRTPLTQSDLSQDSEDVVDHAMKSFTGPITQVPPMYS 121
+ K YEGE+T+G++T+TED SG+ + T Q E VVD + +F G I Q+PPMYS
Sbjct: 63 DYPKAYEGEVTVGFSSTTTEDRSQDTE-TKTIQQPFVEAVVDQVLATFVGEIKQIPPMYS 121

Query: 122 AVKVNKKLYEYARSGEEVERPKRQITISEFRRTSPLYFEKGICRFSFYVSCSKGTYVRT 181
AVKV GK+LYEYAR+G VERP+R +TI R S + +E+G+CRF F VSCSKGTYVRT
Sbjct: 122 AVKVRGKRLYEYARAGITVERPERTVTIFSLERMSDIVYEEGVCRFRFNVSCSKGTYVRT 181

Query: 182 LAVDLGIKLG YASHMSFLKRTSSAGLSITQSLTLEEINEKYKQ-EDFSFLLPIEYGVLDL 240
LAVD+G LGY +HMS L RT S S+ + T E+ E+ +Q E S LLPIE +LD+
Sbjct: 182 LAVDIGKALGYPAHMSDLVRTKSGPFSLEECFTFTELEERLEQGEGSSLLLPITAILDI 241

Query: 241 PKVNLTEEDKVEISYGR-----RILLENEADTLAAFYE 273
P+V + +E + +I +G R + NE L A Y+
Sbjct: 242 PRVQVNKEIEEKIRHGAVLPQKWFNHRFTVYNEEGALLAIYK 284

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5839> which encodes the amino acid sequence <SEQ ID 5840>. Analysis of this protein sequence reveals the following:

Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2698(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 201/295 (68%), Positives = 246/295 (83%), Gaps = 2/295 (0%)

-2123-

Query: 1 MITGIINLKKEAGMTSHDAVFKLRKILHTKKIGHGGTLDPDVVGVLPIAVGKATRVIEYM 60
 MI GIINLKKEAGMTSHDAVFKLRK+L KKIGHGGTLDPDVVGVLPIAVGKATRVIEYM
 Sbjct: 1 MINGIINLKKEAGMTSHDAVFKLRKLLQEKKIGHGGTLDPDVVGVLPIAVGKATRVIEYM 60

5 Query: 61 TESSGIYEGEITLGYATSTEDSSGEVISRTPLTQSDLSQDSEDVVDHAMKSFTGPITQVPPMY 120
 TE+GK+YEG++TLGY+T+TED+SGEV++R+ L + L+E++VD M +F G ITQ PPMY
 Sbjct: 61 TEAGKVYEGQVTLGYSTTTEDASGEVVARSSL-PAVLTEELVDQTMTTFLGKITQTTPMY 119

10 Query: 121 SAVKVNGKKLYEYARSGEEVERPKRQITISEFRRTSPLYF-EKGICRFSFYVSCSKGTYV 179
 SAVKVNG+KLYEYAR+GE VERP+R++TIS F RTSP L F E G+CRFSF V+CSKGTYV
 Sbjct: 120 SAVKVNGRKLIEYARAGESVERPRREVTISLFERTSPLNFTEDGLCRFSFKVACSKGTYV 179

15 Query: 180 RTLAVDLGIKLYASHMSFLKRTSSAGLSITQSLTLEBINEKYKQEDFSFLLPIEYGVLD 239
 RTLAVDLG LG SHMSFL+R++SAGL++ + TL EI + +++ SFLLEPIEYGV D
 Sbjct: 180 RTLAVDLGRALGVESHMSFLQRSASAGLTLETAYTLGEIADMVSKQEMSFLLEPIEYGVAD 239

20 Query: 240 LPKVNLTEDKVEISYGRRIILENEADTLAAFYENRVIAILEKRGNEFKPHKVLL 294
 LPK+ + + + BIS+GRR+ L ++ LAAF+ +VIAILEKR E+KP KVL+
 Sbjct: 240 LPKVIDDTLTELTEISFGRRRLSLPSQEP LLAFAFHGEKVIAILEKRDQEYKPKKVL 294

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1877

A DNA sequence (GBSx1985) was identified in *S.galactiae* <SEQ ID 5841> which encodes the amino acid sequence <SEQ ID 5842>. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2776(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9819> which encodes amino acid sequence <SEQ ID 9820> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12871 GB:Z99109 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 39/145 (26%), Positives = 68/145 (46%), Gaps = 7/145 (4%)

40 Query: 3 MKIRTATLDDSEKLVPLYQELG---YAISLSEIQSILKVILTHSDYGFLLAEDNGKLLA 58
 M I R A D+ + PL+ + A L ++ LK L + + LIAE+NG+ +
 Sbjct: 1 MNIRQAKTSDAAAIAPLFNQYREFYRQASDLQGAFLKARLENHESVILIAEENGFIG 60

45 Query: 59 FVGYHKLYFFFEKSGTYRILALVVNEKHRRKGIASQLINHVQLAKTDGSEVLAINSSLK 118
 F + + Y + L V R KG +L++ K A +G++ L L + +
 Sbjct: 61 FTQLYPTFSSVSMKRIYILNDFVVPHARTKGAGRLLSAAKDYAGQNGAKCLTLQT--E 118

Query: 119 EYRQEAYHFYENLGFKKVSTGFSYY 143
 + ++A YE G+++ TGF +Y
 50 Sbjct: 119 HHNRKARSLEYQNGYEE-DTGFVHY 142

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5843> which encodes the amino acid sequence <SEQ ID 5844>. Analysis of this protein sequence reveals the following:

55 Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0962(Affirmative) < succ>

-2124-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 37/126 (29%), Positives = 64/126 (50%), Gaps = 16/126 (12%)

Query: 18 PLYQE-----LGYAISLSEIQSILKVILTHSDYGFLLIA--EDNGKLLAFVG---YHKLYF 67
 P+ QE LGY +SL ++ +++ + FL +D +LL +V Y LY
 10 Sbjct: 11 PMLQBINAKALGYLVSLDLLERQYERLIEDCHHYFLAYADKDTNQLLGYVHAERYETLY- 69

Query: 68 FEKSGTYRILALVVNEKHRRKGIASQLINHVKQLAKTDGSEVLALNSSLKEYRQEAYHF 127
 + +L L V ++R+GI S L+ ++ A+ +G + LNS+ +R+EA+ F
 15 Sbjct: 70 ---ASDGLNLLGLAVLPAYQRRGIGSALLRALESQARQEGIAFIRLNSA--SHRKEAHAF 124

Query: 128 YENLGF 133
 Y NL +
 15 Sbjct: 125 YRNLDY 130

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1878

A DNA sequence (GBSx1986) was identified in *S.agalactiae* <SEQ ID 5845> which encodes the amino acid sequence <SEQ ID 5846>. Analysis of this protein sequence reveals the following:

25 Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1659(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif 28-30

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:AAF30776 GB:AE002133 conserved hypothetical [Ureaplasma urealyticum]
 Identities = 106/440 (24%), Positives = 206/440 (46%), Gaps = 65/440 (14%)

40 Query: 13 FAINESEYHQLLEQIRGDAFDKEVSEERLEKERLILGEQAKNQLQEVVVE-KDKEIAKLQY 71
 F N+ +Y++L++Q +D LEK+R L E+ KN+ ++ KD + K
 Sbjct: 71 FLANDRDYNELVKQ----RYD-----LEKQDELKEKLKNEGKKAIAHFKDSDEYKNLI 120

45 Query: 72 KVKQFLIEKDNLLKDNEYQLAEQLNOKDMMLRD-----LENQIDRLRLHEHENSLOEA 123
 K ++ + + ++ NE +++ ++ L+ L+N I + ++ +N+ + A
 Sbjct: 121 KAQEKINSLNKTIESNEQSYKKEIENIELKLKSQFDEETKSLKNTIAKQEIKLDNAEKMA 180

50 Query: 124 LTKVERE-----RDAIQNLHIQ-----EKEKDIALASVKSQDY 156
 + + +D I + I+ E +K + + ++S
 Sbjct: 181 IINFKESNEYQKIIKDKIDLDEIEBKLKFAIQAHEDNMKAAKENWESKIVEIKELESKK 240

55 Query: 157 EVQLKAANEQVEFYKNFKAQQSTKAVGESLEHYAETEFNKRHLAFPNAYFEKDNTLSSR 216
 + ++ E +E K K+ + K VGE LE + + +F++ + P+ F K N
 Sbjct: 241 DKEIHKLTESIEQLKREKSS-NVKLVGEELEQWLKNKFDETYSFSCPDMTFTKINEAID- 298

60 Query: 217 GSKGDFIY-----REKDENDLEFL-SIMFEMKNESDDTIKHKHNEFFKELDKDRREKS 269
 G K DF+ +E +D + + S E K E D K KN +K+LD+DR +
 Sbjct: 299 GKKADFLLEFFDFGKEMSNDKLIIFSATIEAKTEFFDNQKGTKNSAHYKKLDQDRINQK 358

60 Query: 270 CEYAVLVTMLEADNDYYNTGIVDVSHKYPKMYVIRPQFFIQLIGILRNAALNTLKYKQEL 329
 EYA+LVT LE ++ + ++ ++Y M+ +RQ+FI L+ ++RN A TLK K
 Sbjct: 359 SEYAILVTELEPEDHF----VIKKINEYKNMFAVRPQYFIPLVDMIRNFA--TLKAKINS 412

-2125-

Query: 330 ALMKEQNIDITHFEEDLDIFKNAFAKN-YNASKNFQKAIDEIDKSIKRMEAV-KAALT 387
 +++ + D EE+LD K N + +K ID+ IK+ E++ ++A
 Sbjct: 413 QIRYE--DRAKIEENLDELKKDIVDNTLKYINDKTKKIIDDSKAIKKAESTIEESAEDI 470

Query: 388 SENQLRLANNKLDDVSVKKL 407
 +L K++++++K+
 Sbjct: 471 INKKLNTLKKKINELTIRKI 490

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5847> which encodes the amino acid sequence <SEQ ID 5848>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3192(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below.

Identities = 310/445 (69%), Positives = 352/445 (78%), Gaps = 22/445 (4%)

- Query: 1 MNEIKCPHCGTAFAINSEYHQLLEQIRGDAFDKEVSEERLEKERLILGEQAKNQLQEVVV 60
 MNEIKCPHC T F INESEY QLLEQ+RG AFD+E+ +RL E +L E+AK+QL EVV
 25 Sbjct: 1 MNEIKCPHCHTLFTINSEYSQLEQVRGQAFDEELKKRLINEIALLEEKAKHQLHEVVA 60
- Query: 61 EKDKETIAKLQYKVKQF-----LIEKDNLL-----KDNEYQLAEQLNQK 98
 +K+ I L +++Q L +KD L+ N +LA QL +K
 30 Sbjct: 61 KKETAITSLTNQLEQIEKEQAYLRQEELAKKDQLIASLEAKLDKLASONALELANQLAEK 120
- Query: 99 DMMLRDLENQIDRLRLEHENSLOEALTKVERERDAIQNLHIOEKEKDLALASVKSDYEV 158
 D + L NQ+D+L LE + + Q L +E+ERD I+NQL +Q KE +L+LASV+SDYE
 35 Sbjct: 121 DKEVVSILTQLDKLALEKDATFQSKLATIEKERDGIKNQALQAKESLSLASVRSDYEA 180
- Query: 159 QLKAAANEQVEFYKNFKAQOSTKAVGESLEHYAETEFNKVRHLAAPPNAYFEKDNLTSSRGS 218
 QLKAAANEQVEFYKNFKAQOSTKA+GESLE YAETEFNKVR APPNA F KDN LSSRGS
 40 Sbjct: 181 QLKAAANEQVEFYKNFKAQOSTKAIGESLELYAETEFNKVRSYAPPNASFVKDNQLSSRGS 240
- Query: 219 KGDFIYREKDNENDLEFLSIMFEMKNESDDTIKKHKNEDFFKELDKDRREKSCEYAVLVIM 278
 KGD+IYRE D N +E LSIMFEMKNE+D T KHKN DFFKELDKDRREK CEYAVLV+M
 45 Sbjct: 241 KGDIYREVDANGVEILSIMFEMKNEADTTKTKHKNSDFFKELDKDRREKDCYAVLVSM 300
- Query: 279 LEADNDYYNTGIVDVSHKYPKMYVIRPQFFIQLIGILRNAALNTLKYQELALMKEQNID 338
 LEADNDYYNTGIVDVSH+Y KMYV+RPQ FIQLIGILRNAALN+L YKQELAL+KEQNID
 50 Sbjct: 301 LEADNDYYNTGIVDVSHYQKMYVVRPQLFIQLIGILRNAALNSLHYQELALVKEQNID 360
- Query: 339 ITHFEEDLDIFKNAFAKNYNSASKNFQKAIDEIDKSIKRMEAVKAALTTSENQLRLANNK 398
 ITHFEEDLD FKNFAKNY SAS NF+KAIDEIDKSIKRME VK LTTSENQLRLANNK
 55 Sbjct: 361 ITHFEEDLDQFKNFAKNYQSASNNFKKAIDEIDKSIKRMEEVKRLTTSENQLRLANNK 420
- Query: 399 LDDVSVKKLTRKNPTMKAKFDALKD 423
 L+DVSVKKLTR+NPIM+ KF+ALKD
 60 Sbjct: 421 LEDVSVKKLTRQNPTMREKFEALKD 445

- 55 SEQ ID 5846 (GBS304) was expressed in *E.coli* as a His-fusion product. The purified protein is shown in Figure 206, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2126-

Example 1879

A DNA sequence (GBSx1987) was identified in *S.agalactiae* <SEQ ID 5849> which encodes the amino acid sequence <SEQ ID 5850>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

```

5   Possible site: 34
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.1845(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5851> which encodes the amino acid sequence <SEQ ID 5852>. Analysis of this protein sequence reveals the following:

```

   Possible site: 51
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
20      bacterial cytoplasm --- Certainty=0.2492(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

25   Identities = 113/180 (62%), Positives = 141/180 (77%)

   Query: 16  LSELVDCFKGKAVPSKAEAGDIRIINLSDMSPLGIDYHNLRTFQDEQSRLLKYLLQEGDV 75
              L +VDCFKGKAV SK  GD+ +INLSDM LGI YH LRTFQ ++R LL+YLL++GDV
   Sbjct: 18  LGTVVDCFKGKAVSSKVVPGDVGLINLSDMGTLGIQYHQLRTFQMDRRQLRLRYLLEDGDV 77

30   Query: 76  LIASKGTVKKVAIFEEQDYPVVASANITILRPTQHIRGYLYKLFFDSEEGQQALENANKG 135
              LIASKGT+KKV +F +Q+ VVAS+NIT+LRP + +RGYY+K F DS  GQ  L+ A+ G
   Sbjct: 78  LIASKGTLKKVCVFHKQNRDVVASSNITVLRPQKLLRGYYIKFFLDSPIGQALLDVADHG 137

35   Query: 136 KAVMNISTKELLNIAIPSIPLFRQDYLIQRYKQGLNDYKRKIARAEQEWERIQNDIRQOL 195
              K V+N+STKELL+I IP IPL +QDYLI Y +GL DY RK+ RAEQEWE IQN+I++ L
   Sbjct: 138 KDVINLSTKELLDIPVPIPLVKQDYLINHYLRGLTDYHRKLNRAEQEWETIQNEIQKGL 197

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1880

A DNA sequence (GBSx1988) was identified in *S.agalactiae* <SEQ ID 5853> which encodes the amino acid sequence <SEQ ID 5854>. Analysis of this protein sequence reveals the following:

```

45   Possible site: 15
   >>> Seems to have a cleavable N-term signal seq.

      INTEGRAL    Likelihood = -7.43    Transmembrane    62 - 78 ( 55 - 82)
      INTEGRAL    Likelihood = -2.87    Transmembrane    130 - 146 ( 130 - 150)
      INTEGRAL    Likelihood = -1.28    Transmembrane    37 - 53 ( 37 - 53)

50   ----- Final Results -----
      bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```


-2127-

A related GBS nucleic acid sequence <SEQ ID 9347> which encodes amino acid sequence <SEQ ID 9348> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:CAA22372 GB:AL034446 putative transmembrane protein
    [Streptomyces coelicolor A3(2)]
    Identities = 38/139 (27%), Positives = 64/139 (45%), Gaps = 5/139 (3%)

    Query: 15 SASVEILCRGWLLPVSATKYSKIVSVSISSIFFGLLHSANNHVSLSIFNLCL-FGLFLS 73
              +A+ E++ RG L +      +++ ++ + FGL+H N +L + + G L+
10  Sbjct: 143 AATEEVVFRGVLFRIEEEHIGTYLALGLTGLVFGMLHLLNEDATLWGALAIIEAGFMLA 202

    Query: 74 LYVILKGNIWGACGIHGAWNCVQGSVFGIEVSGEPMLSNSLVHVKTGYADWISGGKFGVE 133
              N+W G+H WN G VF VSG S L+ G ++GG FG E
15  Sbjct: 203 AAYAATRNLWLTIGVHFGWNFAAGGVFSTVVSNGND-SEGLLDATMSGPKLLITGGDFGPE 261

    Query: 134 GSMIT---SIVLIVACYWL 149
              GS+ + ++L + WL
    Sbjct: 262 GSVYSVGFGVLLTLVFLWL 280

```

20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1881

25 A DNA sequence (GBSx1989) was identified in *S.agalactiae* <SEQ ID 5855> which encodes the amino acid sequence <SEQ ID 5856>, which is a methylase gene homolog. Analysis of this protein sequence reveals the following:

```

    Possible site: 33
    >>> Seems to have no N-terminal signal sequence

30  ----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2192(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35  RGD motif: 264-266

```

A related GBS nucleic acid sequence <SEQ ID 9929> which encodes amino acid sequence <SEQ ID 9930> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

40  >GP:BAA87672 GB:AB016260 Hypothetical gene, methylase gene homolog
    [Agrobacterium tumefaciens]
    Identities = 358/1238 (28%), Positives = 595/1238 (47%), Gaps = 99/1238 (7%)

45  Query: 1072 KEVARIKGMVDIRNAYQEVIQRYDYDKETFNHLLGKLNRTYDSFVKHYGYLNSAV-- 1129
              K V I+ ++ IR+A +EV+ Q      + L +L + SFV+ +G +N
    Sbjct: 497 KHVRIIRKLIPIRDAVREV LKAQEL-----DRPWKDLQVRLRVAWSSFVRDFGPINHTTVS 552

50  Query: 1130 -----NRNLFSDDDKYSLLASLEDESL--DPSGKSVIYTKSLAFEKAL 1170
              N F D L+AS+ED L D + I+T E+ +
    Sbjct: 553 ITEDPESGETRESHRRPNLQPFADDPDCWLVASIEDYDLENDTAKPGAIFT-----ERVVI 607

    Query: 1171 VRPBEKVKVHTALDALNSSLADGRGVDFAYMMSIYQVESQMTLIEELGDLIMPDPPEKYL 1230
              P V + +A DAL L + VD ++ + + ++ ELG I DP
55  Sbjct: 608 SPPAPPV--ITSADALAVVLNERGRVLDLHIAELLHRDPD-DVVAELGSAIFRDP---- 660

```

-2128-

- Query: 1231 NGELTYVSQRQDFLSGDDVTKLEVVDLFDVQDNQDFNWSHYAGLLEAIKPARITLADIDYR 1290
+ ++ +LSG V KL+V + D ++ L ++P + +DI R
Sbjct: 661 -ADGSWQMADAYLSGPVRDKLKVAEAAAALDPV---YNRNV TALAGVQFVDLRPSDITAR 716
- Query: 1291 IGSRWIPLAVYKFAQETFMGKAYELSDQ-EVATVLEVSPIDGVITYQSKFAYTYSNATD 1349
+G+ WIP A F +E MG + E+A+ + G + A T TD
Sbjct: 717 LGAPWIPAADVVAFFVKE-MMGTDIRIHHMPELASWTVEARQLGYLA-----AGTSEWGTD 770
- Query: 1350 RSLGVPASRYDSGRKIFENLLNSNQPTITTKQVVEGDKKKNVTDVEKTTVLRAKETHLQEL 1409
R ++ + LNS P I + +GD ++ V +V T + K +++
Sbjct: 771 RR-----HAGELLSDALNSRVPQIFDTIRDGDSERRVLNVVDTEAAKEKLHKIKDA 821
- Query: 1410 FQGFVAKYFEVQQMIEDTYNRLYNRTVSKSYDGSHTIDGLAQNISLRPHQKNAIQRIVE 1469
FQ ++ P+ + YN +N + + G HL + G + L HQK I RI+
Sbjct: 822 FQRWIWSDPDRTDLRLARVYNDRFNNIAPRKFSGDHLNLPAGSAGFVLYGHQKRGWIWIS 881
- Query: 1470 EKRALLAHEVSGSKTITMLGAGFKLKELMVHKPLYVVPSSLTAQFGQEIIMFFPTKKVY 1529
LAH VG+GKT+TM + + + LG++ K + VVP AQ +E + +PT ++
Sbjct: 882 SGSTYLAHAVGAGKTMMAASIMEQRRLGLIAKAMQVVPGHCLAQAAREFLALYPTARIL 941
- Query: 1530 VTTKKDFAKAKRKQFVSRIITGDYDAIVIGDSQFEKIPMSREKQVITYINDKLEQLREIKL 1589
V + +F+K KR +F+SR T +DAI+I S F I + + I+D+LE + L
Sbjct: 942 VADETNFSKDKRARFLSRAATATWDAAIITHSAFRFIGVPAAFESQMIHDELELYETLLL 1001
- Query: 1590 GSDSDYTV--KEAERSIKGLEHQLEELQKLERDTFIEFENLIDFLFVDEAHFKNIRPI 1647
+ + V K ER +GL+ +LE L +D + +G+D + VDEA F+ +
Sbjct: 1002 KVEDEDRVSRKRLERLKEGLQERLEALST-RKDDLLTIAEIGVDQIIVDEAQEFRKLSFA 1060
- Query: 1648 TGLGNVAGITNTTSKKNVDMEMKVRQVQABHGDRNVVFATGTPVSNISSELTMMDYIQP 1707
T + + G+ S++ D+ +K R ++ + R +V A+GTP++N++ E+F++ +
Sbjct: 1061 TMMSTLKGVDPNQSQRAWDLYVKSRIETINPGRALVLASGTPITNTLGEMFSVQRLMGH 1120
- Query: 1708 DVLERYLVSNFDSWVGAFNGIENSMELAPTGDYQPKKRKKFVNLPELMRIYKETADI- 1766
LE + FD+W FG+ +EL P+G KY+P RF FVN+PEL+ +++ AD+
Sbjct: 1121 AALBERGLHEFDWASTFGDTTTELELOPSG-KYKPVSRFASFVNPELIAMFRSFADV 1179
- Query: 1767 ---QTSMDLDP-VPEAKIIA VESELTQAQKYYLEELVKRSDAIKSGS--VDP SRDNMLK 1820
+ + +P + + V S+ TQA K++ L +R AI+ P D +L
Sbjct: 1180 MPADLREYVKVPAISTGRRQIVTTSKPTQAFKHHQMVLAERIKAEIERERPPQPGDDILLS 1239
- Query: 1821 ITGEARKLAIDMRILIDPTYSLSDNQKILQVVDNVERIYRDGAGDK-----AT 1867
+ + R AID+RL+D + K+ +V N RI++ AG A
Sbjct: 1240 VITDGRHAIDLRLVDADNDNEPDNKLNNLVSNAFRIWKATAGSVYL RHDSKPFVPGAA 1299
- Query: 1868 QMIFSDIGTPK-SKEEGFDVYNELKDLFVDRGIPKEEIAFVHDANTDEKKNLSRKVNNSG 1926
QMIFSD+GT K GF Y ++D + G+P EIAF+ D E K L V +G
Sbjct: 1300 QMIFSDLGITISVEKTRGFSAYRWIRDELIRLGVPAEIAFMQDFKKSEAKQRLFGDVRAG 1359
- Query: 1927 EVRILMASTEKGGTGLNVQSRMKAVHYLDVFWRP SDIVQRNGRLIRQGNMHEVDIYHYI 1986
VR L+ S+E GTG+NVQ R+KA+H+LDVPW PS I QR GR++RQGN H EVDI+ Y
Sbjct: 1360 RVRFLIGSSETMGTVNVQLRLKALHLDVFWLP SQIEQREGRIVRQGNHDEVDIFAYA 1419
- Query: 1987 TKGSFDNYLWQTQENKLYITQIMTSKDPVRS AEDIDE-QTMTASDFKALATGNPYLKLK 2045
T+GS D +WQ E K ++I ++ +R EDI E Q + KA+A+G+ L K
Sbjct: 1420 TEGSLDATMWQNNERKARFIAAALSGDTSIRRLLEDIGEGQANQFAMAKAIASGDQRLMQK 1479
- Query: 2046 MELENELTVLENQKRAFNRSKDEYRHTISYSEKHLPI MEKRLSQYDKDIAQSLATKSQDF 2105
LE ++ LE + A + R + +E+ + + +R+++ +DI + + T +DF
Sbjct: 1480 AGLEADIAERLERLRAAHIDDQHAVRRQLRDAERDIEVSTRRIA EIGQDITRLVPTTGEDF 1539
- Query: 2106 VMRFNDQAMDNR AEAGDYLRK-LITYNRSETKEVRTLASFRGFDLKM-TTRGASEPLPET 2163
M + R EAG L K ++T + + +AS GF+L+ R + T
Sbjct: 1540 TMTVAGKDYSERKEAGRALMKEILT LVQLSPEGEAVIASIGGFELEYHGYRGYKDG YRYT 1599
- Query: 2164 ISLMIVGDNQYTV ALDLK-SDVGTIQRISNAIDHIIDDQEKTELVDKLDKLRVAKVEV 2222
L G + Y + L + + +G + R+ +A+D ++E+ ++ + D + +L +
Sbjct: 1600 TMLKRTGAD-YEIELEPVITPLGAVSRLEHALDDFGERERYRQRLGDARRRLASYQSRG 1658

-2129-

Query: 2223 DKVFPKEEDYQLVKAKYDVLAPLVEKEAEIEEIDAALA 2260
 + + + + L EK ++ E++ ALA
 Sbjct: 1659 E-----GSEFAFAGELAEKHRQLAEVETALA 1684
 Identities = 99/271 (36%), Positives = 153/271 (55%), Gaps = 10/271 (3%)

Query: 607 RDKVETNIVAIRLVKNLEVEHRNASPSEQELLAKYVGWGG--LANEFFD-----DYNPKF 659
 +D+ NI AIRL +E R A+ EQE L ++ G+G LAN F ++ +
 Sbjct: 80 KDRARDNIAAIRLAAEIEASERPATREEQETLIRFTGFGASDLANGVFRFPGELEFRKGW 139

Query: 660 SKEREELKSLVTDKEYSDMKQSSLTAYYTDPSLIRQMWDKLERDGTGGKILDPSMGIGN 719
 + +L+ V + +Y+ + + + A++T ++R +W L+R G+ GG++L+P +GTG
 Sbjct: 140 DEIGSDLEDAVGETDYASLARCTQYAHFTPEFTVRAIWSQLQLGWRGGRVLEPGIGTGL 199

Query: 720 FFAAMPKHLREKSELYGVELDTITGAIKHLHPNSHIEIKGFETVAFNDNSFDLVISNVP 779
 F A MP+ LR+ S + GVELD +T I + L P + I F SFDL I N P
 Sbjct: 200 FPALMPEALRDLSHVTGVELDPVTACIVRLQPRARILTGFARTEL-PASFDLAIGNPP 258

Query: 780 FANIRIADNRYDRP--YMIHDYFVKKSLDLLHDGGQVAIISSGTMDKRTENILODIRET 837
 F++ + +R R +HDYFV +S+DLL G A ++S+GTMDK Q I T
 Sbjct: 259 FSDRTVRSRAYRSLGLRLHDYFVARSIDLLKPGAFAAFVTSSGTMDKADSAARQHIATT 318

Query: 838 TEFLGGVRLPDSAFKAIAGTSVTTDMLFFQK 868
 + + +RLP+ +F+A AGT V D+LFF+K
 Sbjct: 319 ADLIAAIRLPEGSFRADAGTDVVVDILFFRK 349

SEQ ID 5856 (GBS327N) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 148 (lane 8-10; MW 140kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 148 (lane 11-13; MW 115kDa) and in Figure 182 (lane 8; MW 115kDa).

Purified GBS327N-GST is shown in Figure 243, lane 5; Purified GBS327N-His is shown in Figure 235, lane 5.

GBS327C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 148 (lane 14; MW 73kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1882

A DNA sequence (GBSx1990) was identified in *S.agalactiae* <SEQ ID 5857> which encodes the amino acid sequence <SEQ ID 5858>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3656(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2130-

Example 1883

A repeated DNA sequence (GBSx1991) was identified in *S.agalactiae* <SEQ ID 5859> which encodes the amino acid sequence <SEQ ID 5860>. This protein is predicted to be giant membrane protein. Analysis of this protein sequence reveals the following:

5 Possible site: 33
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.3698(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAG19662 GB:AE005054 calcium-binding protein homology; Cbp
 [Halobacterium sp. NRC-1]
 Identities = 22/43 (51%), Positives = 29/43 (67%), Gaps = 1/43 (2%)

 Query: 9 KDSQDQGLTDAQELAL-GTDPQSVDTGQADLEELQSGHSP 50
 +D+D DGL+D E+ + GTDP DTDGDG D EL++G P
 20 Sbjct: 198 RDTDDGSLSDGVEVRVAGTDPTRDTGQDGVDDAAELRAGSLP 240

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1884

A DNA sequence (GBSx1992) was identified in *S.agalactiae* <SEQ ID 5861> which encodes the amino acid sequence <SEQ ID 5862>. Analysis of this protein sequence reveals the following:

 Possible site: 52
 >>> Seems to have no N-terminal signal sequence

30 INTEGRAL Likelihood = -2.39 Transmembrane 1609 -1625 (1609 -1625)
 INTEGRAL Likelihood = -1.81 Transmembrane 30 - 46 (29 - 46)

 ----- Final Results -----
 35 bacterial membrane --- Certainty=0.1956(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif 1600-1604

40 The protein has homology with the following sequences in the GENPEPT database.

 !GB:X57841 antigen I /II [Streptococcus sobrinus] (v...

 >GP:CAA40973 GB:X57841 antigen I /II [Streptococcus sobrinus]
 Identities = 419/1436 (29%), Positives = 608/1436 (42%), Gaps = 310/1436 (21%)

45 Query: 23 KSKKYRTLCSVALGTMVTAVVWGGTVAHADEVTTTSV----DTTIQRTN--NPATNLPEA 76
 K K RTL LGT + A A G A A+E +T+ DT + TE NPATNLP+
 Sbjct: 23 KVKSGRTLSGALLGTAIALSGA--GQKALAEETSTTSTSGGDTAVVGTETGNPATNLPDK 80

50 Query: 77 QPNP-----VSEQTESMASTGQNGAIAVTVPHDVT-----QAVE 112
 Q NP V T + +S VTV D + +
 Sbjct: 81 QDNPSQAETSQAQARQKTGMSVDVSTSELDEAAKSPQEAGVTVSQDATVNGKTVEPSD 140

 Query: 113 EAKAEGVSTVEDSPMDLGNTRSAVET-----NQQIS-----K 144
 EA + +D + + A E NQ+I+ K
 55 Sbjct: 141 EANQKEPEIKDDYSKQAAIDQKATEDYKASVAANQAETDRINQEI AAKKAQYEQDLAANK 200

-2131-

5 Query: 145 AD-----ADTQKQVETINEVTK----TYKADKATYESNKARIEQEN 181
A+ A QK + I + Y A K Y+ AR++ N
Sbjct: 201 AEVERSLMRMRKPRPIYEAKLAQNQKDLAAIQANSQSAAYAAAKEAYDKEWARVQAAN 260

10 Query: 182 KELSQAQYEGANQTGKETNAWVDTKVNDLKARYADADVTVKEQ-----VVSSGNGTSVL 234
+AYE A N + ++ ++ R A AD K +GN +
Sbjct: 261 AAAKAYEEALAANTAKNDQIKAEIEAIQORSADKADYEAQKLAQYKDLAAQAAGNAANE 320

15 Query: 235 DY----TNYGKAVETIQSTNEQAVADY----LTKKTKADDIVAKNQAIQKENE 280
DY Y + + +Q+ N A Y K I A+N+AIQ+ +A
Sbjct: 321 DYQAKKAAEQELARVQAANAAKQAYEQALAANSAKNAQITAENEAIQQAQAKADYEA 380

20 Query: 281 -----GLANAKADNEAIERRNQAGQAAVDAEN--RAGQAAVDQANQEKQQLVSDRAA 330
LA A++ N A E Q AA + E +A AA QA +++ Q + + A
Sbjct: 381 KLAQYQKDLAAQSGNAANEADYQEKLAAYEKELARVQAANAAKQAYEQVQQAANAKNA 440

25 Query: 331 EIEAITKRNKEKEAAARKENEIDAINTKEMERYQORDLAEIS----- 372
EI + +E+ A A+ + E + +E+ +Y++DLAE
Sbjct: 441 EITEANRAIRERNAKAKTDYELKLSKYQEELAQYKDLAEYPAKLQAYQDEQAIAKALA 500

30 Query: 373 -----KGEEGYISEALAQALNLNNGEPQAOHGAI TRN----- 404
K E+G +SE AQ+L + + EP AQ +T
Sbjct: 501 ELEKHKNEDGNLSEPSAQSL-VYDLEPNAQVALVTDGKLLKASALDEAFSHDEKNYNNHL 559

35 Query: 405 --PDQI-----ISTGDALLGGYSRIIDSTGF-----FVYDMFKTGETLS 441
PD + +++ L G + D G+ F + K G++ +
Sbjct: 560 LQPDNLNVTYLEQADDVASSVELFGNFG---DKAGWTTVSNGAEVKFASVLLKRGQSAT 616

40 Query: 442 FNYQNLQHARFDGKKISRVTYDITNLVSPAG-----TNAVKLVPNDPTEGFIAYRNDGN 496
Y NL+++ ++GKKIS+V Y T V P T V L + DPT G A G
Sbjct: 617 ATYTNLKNISYNGKKISKVVKYT--VDPSKFQNPNTGNVWLGIFTDPTLGVFASAYTGQ 674

45 Query: 497 GDWRD-----KMEFRVAKYYLEDGSQVTFSEKPGVFTHSSLNHNIDIGLEYVKDSSGKFV 553
+ T K EF +Y EDG+ + F + + +SLN +E KD SG FV
Sbjct: 675 NEKDTISIFIKNEF----TFYDEEDGNPIDFDN---ALLSVASLNREHNSIEMAKDYSGTFV 727

50 Query: 554 PINGSTVQVTN-----EGLARSLGSNRASDLNLPEEWDTTSSRYAYKGAIV 599
I+GS++ N EG + RAS+ WD+ + ++ GA
Sbjct: 728 KISGSSIGENKNGMIYATDTLNFKKGEGGSLHTMYTRASEPG--SGWDSADAPNSWYGAGA 785

55 Query: 600 STVTSQNTY-----TVTFGQGDMPQNVGL-----SYWFALN----- 630
++ N Y T +MPQ G + W++LN
Sbjct: 786 VRMSGPNNYITLGATSATNVLSLAEMPQVPGKDNNTAGKKPNIWYSLNGKIRAVNVPKVTK 845

60 Query: 631 --TLPVARTVTPYSPKPHVTVEL-----EPIPEPITVTPDIYTPKFTTPEKPVFTT---- 679
P P P V EL EP EP TP P PEKPV T
Sbjct: 846 EKPTPFVEPTKPDPTYEVEKELVDLPVEPKYEP-EPTPPSKNPDQSIPEKPVVEPTYEVE 904

65 Query: 680 ----PKPLDEVVQPSLTILTQVT-----LPVKPIPKELTPP-----QVPTV 716
P P++ + T + T PV+P + LPTPP VPTV
Sbjct: 905 KELEPAPVEPSYEKEPTPPQSTPDQEEPEKPVPEPSYQSLPTPPVEPVYETVPGPVSVPVTV 964

70 Query: 717 HYHAYRLTTTSEIMKEVVNSDQANLHEKTVAKDSTVIYPLTVDALSPNRAQTSLIFEDY 776
YH Y+L + KE+ N D ++ + VAK STV + L L R +TTS + D
Sbjct: 965 RYHYKLAQVQGVTKYKIKNODDLIDKTLVAKQSTVKFQLKTADLPAGRPEPSTSFVIMDP 1024

75 Query: 777 LPAGYLFDKETTQKENGNYVLSFDETGNFVTLTAKENLLQEVNKDLTQVYQLTAPKLYGS 836
LP+GY + E T+ + + S+D + VT TA L +N+DLT+ P + G
Sbjct: 1025 LPSGYQLNLEATKVASPGFEASYDAMTHTVTFTATAETLAALNQDLTKAVATYPTVVGQ 1084

80 Query: 837 VQNDGATYSNSYKLLLNKGTNNAYTVTSNVVTVRTPG----DGETTTLITPDKNENAD 891
V NDGATY+N++ L++N +AY + SN+V V TPG D + ITP K N+N +
Sbjct: 1085 VLNDGATYTNFTLMVN---DAYGIKSNIVRVTTGKPNPDNPSNNYITPHKVNKNEN 1140

85 Query: 892 GVLINDTVVALGTTNHRYRLTWDLQYKGDRAKETIARGFFVDDYPEEVLDVVENGTAI 951
GV+I+ V GTTN+Y LTWDLQYKGD+SAKE I +GFF+VDDYPEE LD+ + +
Sbjct: 1141 GVVIDGKSVLAGTTNYELTWDLQYKGDKSAKEIIQKGFFVDDYPEEALDLRTDLIKL 1200

-2132-

Query: 952 TTLDGQKVSIGITVKNYASLNEAPKDLQDKLAKAKITPTGAFQVFMPPDDNQAFYDQYVQTG 1011
 T +G+ V+G++V +YASL AP +QD L +A I P GAFQVF DD QAFYD YV TG
 Sbjct: 1201 TDANGKAVTGVSVADYASLEAAPAAVQDMLKKANIIIPKGAFQVFTADDPQAFYDAYVVTG 1260

Query: 1012 TSLALLTKMTVKDSLYGQTKTYTNKAYQVDFGNGYETKEVTNTLVSPEPKKQ-NLNKDKV 1070
 T L ++T MTVK + +Y N+AYQ+DFGNGYE+ V N + P+K L D
 Sbjct: 1261 TDLTIVTPMTVKAEMGKTGGSYENRAYQIDFGNGYESNLVNNVVKINPEKDVILTMDPA 1320

Query: 1071 D---INGKPMVLGTQNHYYTLSDWLDQYRGIKADNSQIAQGFFVDDYPE-----EALLPD 1122
 D ++G+ + + +Y L + I AD+++ + F DDY +
 Sbjct: 1321 DSTNVDGQTIALNQVFNYRLIGGI-----IPADHAEELFEYSFSDDYDQTDGQYTGQYKA 1375

Query: 1123 EAAIQFVTSDGKTV-SGITVKS--SOLLEAPKTLQAAFSKQKIQPKGAFQVFMPE 1175
 A + DG + +G + SY +Q+ EA + F + ++ F E
 Sbjct: 1376 FAKVDLTLKDGTTIKAGTDLTSYTEAQVDEANGQIVVTFKEDFLRSVSVDSAFQAE 1431
 Identities = 209/442 (47%), Positives = 280/442 (63%), Gaps = 27/442 (6%)

Query: 1198 TVLETMLNSGKSY-ENVAYQVDFGQAYETNTVTNFVPK-----VTPHKSNTNQ 1244
 TV+ +LN G +Y N V+ ++N V P +TPHK N N+
 Sbjct: 1080 TVVGQVLNDGATYTNFTLMVNDAYGIKSNIVRVTTGKPNPDNPSNNYITPHKVNKNE 1139

Query: 1245 EGISIDGKTVLPNTVNYKIVLDYSQYKDMVVTDDVLAKGYMVDYDYPEEALTLNPDGIQ 1304
 G+ IDGK+VL T NYY++ D QYK +++ KGF+ VDDYPEEAL L D I+
 Sbjct: 1140 NGVVIDGKSVLAGTTNYYELTWDLQYKGDKSAKEIIQKGFYVDDYPEEALDLRTDLIK 1199

Query: 1305 VLDKDGNRVSGISVSTYASLSEAPKVVQDAMAKRQFTPKGAIQVLSDDPKVFYDITYVKT 1364
 + D +G V+G+SV+ YASL AP VQD + K PKGA QV ++DDP+ FYD YV T
 Sbjct: 1200 LTDANGKAVTGVSVADYASLEAAPAAVQDMLKKANIIIPKGAFQVFTADDPQAFYDAYVVT 1259

Query: 1365 GQTLVVTLPMTVKNELTKTGGQYENTAYQIDFGLAYVTETVNNVVKLPQKDVVIDLSH 1424
 G L + PMTVK E+ KTG G YEN AYQIDFG Y + VNNVVK++P+KDV + +
 Sbjct: 1260 GTDLTIVTPMTVKAEMGKTGGSYENRAYQIDFGNGYESNLVNNVVKINPEKDVILTMDP 1319

Query: 1425 KDA-SLDGKEVALHQTFNYRLVGAMIPSNRATDLFEYGFEDNYDEKHDEYNGVRSYLMT 1483
 D+ ++DG+ +AL+Q FNYRL+G +IP++ A +LFEY F D+YD+ D+Y G Y+++
 Sbjct: 1320 ADSTNVDGQTIALNQVFNYRLIGGIIPADHAEELFEYSFSDDYDQTDGQYTGQYKAFKV 1379

Query: 1484 DVILKDGSVLKEGTEVTKYTLQQVDITENGLVSI SFDKSFLETVSDDSAFQADVYLQMKRI 1543
 D+ LKDG+++K GT++T YT QVD NG + ++F + FL +VS DSAFQA+VYLQMKRI
 Sbjct: 1380 DLTLDGTTIKAGTDLTSYTEAQVDEANGQIVVTFKEDFLRSVSVDSAFQAEVYLQMKRI 1439

Query: 1544 AAGQVENTYLHTVNGYVISSNTVTHTPQPEEPSNPQ-----TPQPIETIEPPV 1595
 A G NTY++TVNG SSNTV T TP+P++PSP P P Q PP
 Sbjct: 1440 AVGTANTYVNTVNGITYSSNTVTRSTPEPKQSPVDPKTTTTTVVFQPROGKAYQPAPPA 1499

Query: 1596 PASILPNTGEQES----LLGLI 1613
 A LP TG+ + LLGL+
 Sbjct: 1500 GAQ-LPATGDSSNAYLPLILGV 1520
 Identities = 100/210 (47%), Positives = 137/210 (64%), Gaps = 4/210 (1%)

Query: 1060 PKKQNLNKDKVDINGKPMVLGTQNHYYTLSDWLDQYRGIKADNSQIAQGFFVDDYPEEAL 1119
 P K N N++ V I+GK +L GT N+Y L+WDLQY+G K+ I +GF++VDDYPEEAL
 Sbjct: 1132 PHKVNKNENGVIDGKSVLAGTTNYYELTWDLQYKGDKSAKEIIQKGFYVDDYPEEAL 1191

Query: 1120 LPDEAAIQFVTSDGKTVSGITVKSYSQLEAPKTLQAAFSKQKIQPKGAFQVFMPEPQA 1179
 I+ ++GK V+G++V Y+ L AP +Q K I PKGAFQVF +DPQA
 Sbjct: 1192 DLRTDLIKLTDANGKAVTGVSVADYASLEAAPAAVQDMLKKANIIIPKGAFQVFTADDPQA 1251

Query: 1180 FFESYVTKGENITIVTPMTVLETMLNSGKSYENVAYQVDFGQAYETNTVTNFVVKVTPHK 1239
 F+++YV G ++TIVTPMTV M +G SYEN AYQ+DFG YE+N V N VPK+ P K
 Sbjct: 1252 FYDAYVVTGTDLTIVTPMTVKAEMGKTGGSYENRAYQIDFGNGYESNLVNNVVKINPEK 1311

Query: 1240 SNT----NQEGISIDGKTVLPNTVNYKIV 1265
 T + ++DG+T+ N V Y+++
 Sbjct: 1312 DVTLTMDPADSTNVDGQTIALNQVFNYRLI 1341

There is also homology to SEQ ID 598.

-2133-

SEQ ID 5862 (GBS76) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 2; MW 17.4kDa). The GBS76-His fusion product was purified (Figure 196, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 294), which confirmed that the protein is immunoaccessible on GBS bacteria.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1885

- 10 A DNA sequence (GBSx1993) was identified in *S.agalactiae* <SEQ ID 5863> which encodes the amino acid sequence <SEQ ID 5864>. This protein is predicted to be abortive infection bacteriophage resistance protein (abiEi). Analysis of this protein sequence reveals the following:

Possible site: 31
>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2765(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 A related GBS nucleic acid sequence <SEQ ID 9931> which encodes amino acid sequence <SEQ ID 9932> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

- >GP:AAB52382 GB:U36837 AbiEi [Lactococcus lactis]
Identities = 51/206 (24%), Positives = 90/206 (42%), Gaps = 23/206 (11%)
- 25 Query: 17 KNNGIVTNKDCALGIPTIYLTRLEKEGIIIFRVEKGIFLTONGDYDEYFFQYRFPKAIF 76
K G + K + GI YL + + + V+KG+++ + D + FQ ++ KA+
Sbjct: 76 KYKGNIIIRKIVRDEGISDYLRKFLVLYNLTETVDKGVYIFPHKKKDSLFIQKQYKSKAVI 135
- 30 Query: 77 SYISALYLQQFTDEIPQYFDVTVPGRYF-----NTPPANLNI 114
S+ ++LVLQ D IPQ ++VP Y N N+ I
Sbjct: 136 SHETSLYLQDVIDYIPQKIQMSVPEKYNISRIQEPHENRLTSYNYVDINNNIMDKNIPI 195
- 35 Query: 115 HFV-SKEYSELGTTTVPMPGNNVRVYDFERIICDFVIHREKIDSELVFKTLQSYGNYPK 173
+ V +K S + TV + +G +RV R I D + K + E+ + ++ Y
Sbjct: 196 NLVRNKSISFTQIETVNSFLGLPLRVTSIARSIVDVLKPSHKAEEEVKEQAIIKYLERFP 255
- 40 Query: 174 KNLAKLYEYATKMTLEKVKQTLEVL 199
N+ +L A N L++++ L +L
Sbjct: 256 DNIVRLKRIAKTONVLKELEYLILL 281

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1886

- 45 A DNA sequence (GBSx1994) was identified in *S.agalactiae* <SEQ ID 5865> which encodes the amino acid sequence <SEQ ID 5866>. This protein is predicted to be abortive infection bacteriophage resistance protein (abiEii). Analysis of this protein sequence reveals the following:

Possible site: 43
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.12 Transmembrane 260 - 276 (259 - 277)

50

-2134-

----- Final Results -----

bacterial membrane --- Certainty=0.1447(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB52383 GB:U36837 AbiEii [Lactococcus lactis]
 Identities = 76/276 (27%), Positives = 135/276 (48%), Gaps = 19/276 (6%)

Query: 14 SKNTGLTFNSVMTYYFLEVILKKLSQSSYSNHYIFKGGFLLSNVIGVESRSTVDIDFLFH 73
 ++N + + Y E L +LS S Y ++ KGGFL+ + R+T D+D

Sbjct: 12 TRNDDIGIENYRIRYATERFLRLSASQYKEKFVLKGGFLIGVTYNLSQRTTKDLDTALI 71

Query: 74 QITLSEETVKQQLKEIL-ADSEEGISFVFIQSITTIKESDDYGGYRATISCQLE--NIKQV 130
 +++++ + EI D E+ + F ++ +T+ ++ Y GYRA + N +

Sbjct: 72 DFKSDAQSIERVITEICNIDLEDQVLFKLKELTSSQDMRIYPGYRAKLKMMFPDGNTRID 131

Query: 131 IHLDIATGDVVTQPITYDYKAIFDE----DNFPITAYTITETILAEKLQTIYSRNFLNS 185
 LDI GD +TP+ IF+E ++AY ETI AEKL+TI +R +N+

Sbjct: 132 FDLDIGVGRITPEAKKIKIPLIFNEVKGVEKQIEVLAYPKETIQAEKLETILTRGKVNT 191

Query: 186 RSKDFYDVYIL--SKLKKKDIDFNQLKNACQRTFSYRE-TELDFEKIIE-----LLERFK 237
 R KD+YD ++L + I F A + T+ +R T+ E++ E L E +

Sbjct: 192 RMKDYVDFHLLLTQDENSNSISFYF--AFKNTWBFNRNPTQFIDEELFEDWLFILDEILE 248

Query: 238 SDPTQNQQWQNYSKKYSYTKGISLANVLDEMISLIT 273
 S + + W NY K +Y K +++ ++ E+ ++

Sbjct: 249 SKELKEKYWPNIYKDRNYAKHLNMDDIISEIKEFVS 284

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1887

A DNA sequence (GBSx1995) was identified in *S.agalactiae* <SEQ ID 5867> which encodes the amino acid sequence <SEQ ID 5868>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1137(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1888

A DNA sequence (GBSx1996) was identified in *S.agalactiae* <SEQ ID 5869> which encodes the amino acid sequence <SEQ ID 5870>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

-2135-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2782(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 10 vaccines or diagnostics.

Example 1889

A DNA sequence (GBSx1997) was identified in *S.agalactiae* <SEQ ID 5871> which encodes the amino acid sequence <SEQ ID 5872>. Analysis of this protein sequence reveals the following:

15 Possible site: 21
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood =-10.14 Transmembrane 310 - 326 (301 - 334)
 ----- Final Results -----
 20 bacterial membrane --- Certainty=0.5055(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:AAG38044 GB:AF295925 Orf28 [Streptococcus pneumoniae]
 Identities = 272/344 (79%), Positives = 307/344 (89%)
 Query: 568 VYVNPAPYFPPKVIQVQTTLPTIGQPGGDEFERAKAIYDYLKSKGATNQAI AAILGNWSV 627
 +YVNP FYFPPKVIQ+QTTILP IGQPGGDEFERAK IY++LKS+GA+ QAIAAILGNWSV
 30 Sbjct: 1 MYVNPQFYFPPKVIQLQTTILPAIGQPGGDEFERAKHIYEFLKSCGASPQAIAAILGNWSV 60
 Query: 628 ESSINPKRAEGDYLSPVPGATDSSWDDEGWLTLNGPTIYNGRYPNILKRGGLGQWTDTA 687
 ESSINPKRAEGDYL+PPVG WDDE WL + GP IY+G YPNIL RGLGLGQWTDTA
 35 Sbjct: 61 ESSINPKRAEGDYLTPVGVPIPPWDDSWLAIGGPAIYSGAYPNILHRGLGQWTDTA 120
 Query: 688 DGSRRHTLLLEYAKGKHQKWDYDLGLQDLDFMLYGDSPYYTNWLKDDFFKNSGSPASLAQLFL 747
 DGS RHT LL YA+ +++KWDYDL LQDLDFML+GDSPYY +WLKDDFFKN+GS A+LAQLFL
 40 Sbjct: 121 DGSTRHTALLNYARTQNKWYDLDLQDLDFMLHGDSPYYQSWLKDDFFKNTGSAANLAQLFL 180
 Query: 748 IYWEGNSGDKLLERQTRASEWYYQIEKGFSQPNNGGTAQSDPKALEAVREDLFENSIPGGG 807
 YWEGNSGDKLLERQTRA+EWYYQIEKGFSQ NGG A+SDP++LE VR DL+++S+PGGG
 45 Sbjct: 181 TYWEGNSGDKLLERQTRATEWYYQIEKGFSQTNNGQAQSDPQSLGVRGDLVDHSPVPGGG 240
 Query: 808 DGMGYAYGQCTWGVAAARINQLGLKLGKNGEKIPIISTMGNGQDWVRTAASLGGETGTSP 867
 DGM YAYGQCTWGVAAAR+NQLGLKLG+NGEKI II+TMNGNQDWV T++SLGGETG++P
 50 Sbjct: 241 DGMAYAYGQCTWGVAAARMNQLGLKLGKNGEKIISTINTMGNGQDWVATSSSLGGETGTSP 300
 Query: 868 QEGAILSFAGGGHGTPTFYGHVAFVEKVYPDGSFLISETNYNGN 911
 + GAI+SF GG HGTP YGHVAFVEKVY DGSFL+SETNY GN
 Sbjct: 301 RAGAIVSFVGGTHGTPASYGHVAFVEKVYDDGSFLVSETNYNGN 344

SEQ ID 5872 (GBS74d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 121 (lane 3 & 4; MW 95.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 121 (lane 5-7; MW 70.5kDa) and in Figure 179 (lane 9; MW 70.5kDa).

55 GBS74d-His was purified as shown in Figure 233, lane 7-8.

-2136-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1890

A DNA sequence (GBSx1998) was identified in *S.agalactiae* <SEQ ID 5873> which encodes the amino acid sequence <SEQ ID 5874>. This protein is predicted to be TrsE-like protein. Analysis of this protein sequence reveals the following:

Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.5526(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG38042 GB:AF295925 Orf26 [Streptococcus pneumoniae]
Identities = 618/782 (79%), Positives = 712/782 (91%), Gaps = 1/782 (0%)

Query: 1 MKKLKHSMSK-TSSNDKKQKTKTQKQEISPSTVNTLAYQGLFQNGLMQVSPSYFSQTYL 59
MK+ +++K + TS+ +KK++ K +K+E+ PST NTL+YQ L+QNGLMQV YFSQ+YL
Sbjct: 3 MKRKSNTLKKQQTSTTNKKEEVKDKKEEVLPTANTLSYQALYQNGLMQVKEDYFSQSYL 62

Query: 60 LGDVNYQTVGLDDKGAIVEKYSDLINSLDDKTNFQLTIFNQKVNLEKFRKSILYPLQEDG 119
LGDVNYQTVGL+DKGAI+EKYSDLI SLDD+TNFQLTIFN+++NLEKFR S+LY +EDG
Sbjct: 63 LGDVNYQTVGLEDKGAIIEKYSDLIKSLDDQTNFQLTIFNKRNLNLEKFRHSVLYEEKEDG 122

Query: 120 FDTYRDELNRMDANLEAGENNFSAVKFLSFGKSDQTPKLAFRSLSQIGEYFKSGFSEID 179
+D+YR ELNRMM+ NL++GENNFSAVK +SFG+ D PK A+RSLSQIGEYFKSGFSEID
Sbjct: 123 YDSYRKELNRMMNQNLDSGENNFSAVKLISFGRKDSNPKQAYRSLSQIGEYFKSGFSEID 182

Query: 180 VSLGLLGGEERVNLADMLRGENHLPFSYKDLTSLGQSTKHFIAPTYLSFKHKNHIELD 239
L GEERVN+LADMLRGE+HLPFSY+DLT SQQ+T+HFIAP L FK+KN++++D
Sbjct: 183 ARFESLAGEERVNLADMLRGEHLPFSYRDLTRSGQTTRHFIAPNLDFKNKNYLQIND 242

Query: 240 RLLQIVYVRDYGMEIGDKFIRDLMQSDLEVMISLHAKGSTKSETMTKLRTKKTLMESQKI 299
RLLQIVYVRDYGMEIGD+FIRDLMQ DLE+++SLHA+ STKS+ M KLRTKKTLMESQKI
Sbjct: 243 RLLQIVYVRDYGMEIGDQFIRDLMQGDLELIVSLHAQSSTKSDAMKKLRTKKTLMESQKI 302

Query: 300 GEQQKMARTGIYLEKVGHVLENNIDEAEALLQTMTQTGDKLFDTVFLIGVLADTEDQLKQ 359
GEQQK+ARTGIYLEKVGHVLE+NIDEAE LL+TMT+TGDKLF TVFLIGV E++LKQ
Sbjct: 303 GEQQKLARTGIYLEKVGHVLESNIDEAEELLKTMETGDKLFQTVFLIGVFGQDEEELKQ 362

Query: 360 SLDIIQVAGSNDMIIDNLTYMQEAAFNSLLPFGKNYLEGVSRSLTNSIAVNAPWTSVD 419
+LD ++QVAGSND++ID L YMQEAAFNSLLPFG ++LEGVSRSLTNSIAVN+PWTSD
Sbjct: 363 ALDTVQVAGSNDLMIDKLPYMQEAAFNSLLPFGCDFLEGVSRSLTNSIAVNSPWTSD 422

Query: 420 IHDKGGKFYGINQISSNIIIDRGKLNTPSGLILGTSGAGKGMATKHEIISTKLKEADSD 479
+ D+ GK+YGINQISSNII+IDR LNTPSGLILGTSGAGKGMATKHEII+TK+KE+ +
Sbjct: 423 LQDRSGKYGINQISSNIIITIDRSLNTPSGLILGTSGAGKGMATKHEIITTKIKESGEN 482

Query: 480 TEIIIIVDPENEYSIIGQAFGGESIDIAPDSTTFLNVLELSDENMDEDPVKVSEFLLSWI 539
TEIIIIVDPE EYS+IG+ FGGE IDIAPDS T+LNVL+LS+ENMDEDPVKVSEFLLS+I
Sbjct: 483 TEIIIIVDPEAEYSVIGRTFGGEMIDIAPDSETYLNVLDLSEENMDEDPVKVSEFLLSFI 542

Query: 540 GKLLDRKMDGREKSLIDRVTRLTYKHFDTPSLVEWVFLSQQPEQEAQDLALDMELYVEG 599
GKLLDRKMDGREKS+IDRVTRLTY+ F PSL EWVFLVLSQQPE+EA++LALDMELYVEG
Sbjct: 543 GKLLDRKMDGREKSIIDRVTRLTYQSFKEPSLEEWVFLVLSQQPEEEAQNALDMELYVEG 602

Query: 600 SLDIFSHRTNIKTDSHFLIYNVKKLGDELKQIALMVIFDQIWNRVVKNQKLGKKTWIIYFD 659
SLDIFSH+TNI+T S+FLIYNVKKLGDELKQIALMV+FDQIWNRVV+NQKLGKKTWIIYFD
Sbjct: 603 SLDIFSHKTNIQTGSNFLIYNVKKLGDELKQIALMVVFDQIWNRVVRNQLGKKTWIIYFD 662

```

Query: 660 EMQLLLLDKYASDFFFKLWSRVRKYGA IPTGITQNVETLLLDANGRRRIANSEFMILLKQ 719
      E++LLLLDKY SDDFFFKLWSRVRKYGA PTGITQNVETLLLD NGRRIIANSEFMILLKQ
Sbjct: 663 EIELLLLDKYPSDFFFKLWSRVRKYGASPTGITQNVETLLLDPNGRRIIANSEFMILLKQ 722

5
Query: 720 AKSDREELVHMLGLSKELEKYLVNPEKGAGLIKAGSTVVPFNKIPQHTKLFDMISTDPE 779
      AK+DREELV +LGLSKELEKYLVNPEKGAGLIKAGS VVPFNKIPQ ++LFDIM +DP+
Sbjct: 723 AKNDREELVQLLGLSKELEKYLVNPEKGAGLIKAGSVVVPFNKIPQGSQFLDIMRSDPD 782

10
Query: 780 KM 781
      KM
Sbjct: 783 KM 784

```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

```

20 Lipop: Possible site: -1      Crend: 10
    McG: Discrim Score:      -26.26
    GvH: Signal Score (-7.5): -3.87
        Possible site: 55
    >>> Seems to have no N-terminal signal sequence
    ALOM program   count: 0 value:   6.26 threshold:   0.0
        PERIPHERAL Likelihood =   6.26      335
25     modified ALOM score:  -1.75

    *** Reasoning Step: 3

    ----- Final Results -----
30     bacterial cytoplasm --- Certainty=0.5526(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial outside --- Certainty=0.0000(Not Clear)

```

-2138-

```

      130      140      150      160      170      180      190
894      921      951      981      1011      1041      1071      1101
5  ERVNVVLADMLRGENHL-PFSYKDLTSLGQSTKHFIAPTYLSFKHKNHIELDDRLQLIVYVRDYGMELGDKFIRDLMQSDL
   ||::| : : : |::| || || ||::| : : : : || : |:::
ERLNLHGVYHPDGEIFNFDWKWLAPSGSLSTKDFIAPSSLCFGNAKTFGMGKYGAVSFLQILSPELSDMLADFLNTES
      210      220      230      240      250      260      270

1131      1161      1191      1221      1251      1281      1311      1341
10 EVMISLHAKGSTKSETMTKLRTKKTLMESQKIGEQQKMARITGIYLEKVGHVLENNIDEAEALLQMTQTGDKLFDTVFLI
   |::|| : : : : |::| || || ||::| : : : |::| : : || ||:
GVLVNLHVQAIEQTAKITIKRKITDLDAMKIAEQKKAVERSGYDMDILPSDLATYGEDAKLLTKLQTRNERLFLQTLFLV
      290      300      310      320      330      340      350

1371      1401      1431      1461      1491      1521      1551
15 GVLADTEDQLKQSLDIKQVAGSNDMIIDNLTVMQEAFAFNSLLPFGKNYLEGVSRSLTNSNIAVNAPWTSVDIHDKGGK-
   :||::| : : || : : | || : | ||: | : || ||::| | : : : |
LNVAETKQKLNNDVFQAAGVAQKHNCPLVRLDYQQEQGLASSPLGVNQI-KIQRSLTSSVAVFVPFVTQELFQGGGAAM
      370      380      390      400      410      420      430

1608      1638      1668      1698      1728      1758      1788      1818
20 FYGINQISSNIISIDRGKLNTPSGLLIGTSGAGKGMATKHEIISTKLKEADSDTEIIVDPENEYSIIGQAFGGESIDIA
   :||| | |::| : : | : || |::| : | ||: | : | || || : : : | : :
YYGINAKSRNMIMLDRKQARCPNALKLGTPGSGKSMSCSEIVSVFLTPD---DIFISDPEAEYPLVKRLHGVIRLS
      450      460      470      480      490      500      510

1848      1875      1905      1935      1959      1989      2019
25 PDSTTFLNVLLELS-DENMDEDPVKVKSEFLLSWIGKLLDRK--MDGREKSLIDRVTRLTYKHFDTPSLVEWVFLS----
   | | |::| : : : |::| : |::| : : | : |::| | : : : | : : |
PTSKDFVNPLDINLNYSEDDNPLALKSDFVLSFCELVMGGKNGLEAIEKTVIDRAVRVIYRPLYADPRPENMPILSDLHK
      530      540      550      560      570      580      590

2058      2088      2118      2148      2178      2208      2238      2268
35 ---QQPEQEAQDLALDMELYVEGSLDIFSHRTNIKTDSHFLIYNVKLGDELKQIALMVIQIWNRVVKNQKLGKKTWI
   | || : | : || | ||::| : ||: : : : |::| : ||: : || | | | : || |
ALLDQHVPEDRVAQALDLYVSGSLNVFNHRTNVDIGNRLVSFDIKELGKQLKGLMLIVQDQIWRVTANRSQGKATWY
      610      620      630      640      650      660      670

2298      2328      2358      2388      2418      2448      2478      2508
40 YFDEMQLLLLDKYASDFFFKLWSRVKYGAIPTGITQNVETLLLDANGRRITANSEFMILLKQAKSDREELVHMLGLSKE
   : || : || : : : : | || | || | ||: || : ||::| : || | ||: | || |
FADEFHLLKKEEQTAAYSAEIWKRFKRWGGIPTGATQNVKDLLSSPEIENILBNSDFITLLNQASGDRKILAEERLNSTE
      690      700      710      720      730      740      750

2538      2568      2598      2628      2658      2688      2718      2748
45 LEKYLVNPEKGAGLIKAGSTVVPFNKIPQHTKLFDMISTDPEKMRT*DERG*KASQTG*AKLSKQLKISSYALSERS*D
   : ||: | | | | : : |::| | ||::| : |::| :
QQKYIDNSEPGGLLIFENVVLPFTNPPIPHNTQLYKIMTIRLNEVAGV
      770      780      790

```

A related GBS gene <SEQ ID 8927> and protein <SEQ ID 8928> were also identified. Analysis of this protein sequence reveals the following:

This protein might be involved in vancomycin resistance

The protein has homology with the following sequences in the databases:

```

55 >GP|8100663|gb|AAF72347.1|AF192329_8|AF192329 TrsE-like protein
   {Enterococcus faecalis}

```

Score = 427 bits (1086), Expect = e-118

Identities = 257/785 (32%), Positives = 431/785 (54%), Gaps = 28/785 (3%)

```

60 Query: 9  DKKQKTKTKQKQIS-----PSTVN-TLAYQGLFQNGLMQVSPSYFSQTYLLGDV 56
   +K + T+ Q++EI P T ++ Y+ ++ +G+ +VSP FS+ D+
Sbjct: 11 EKTKLTRAQRKEIDAVIRKYKGDGRPHTAQQQSIPYEVMPDGVCRVSPGVFSKCIEFADI 70

```

```

65 Query: 57 NYQTVGLDDKGAIVEKYSIDLINSLDDKTNFQLTIFNQKVNLEKFRKSILYPLQEDGFDTY 116

```

-2139-

+YQ D + AI EK DL N +D + Q + N+KV+ ++ KS Q D FD
 Sbjct: 71 SYQLAQPDQTQTAIFEKLCALYNYVDASIHQFSFLNRKVDVQYAKSFEIAPQGDDFDDI 130
 Query: 117 RDELNRMDANLEAGENNFSVAVKFLSFGKSDQTPKLAFRSLSQIGYFKSGFSEIDVSLG 176
 R E ++ L G N K+L+F ++ K A L +IG F +
 Sbjct: 131 RAEYTGILQQLANGNNGMVKTLYLTFTIEAESVKAARARLKRIGFDLLGYFKSMGAVAH 190
 Query: 177 LLGGEEVRVNLADMLRGENHL-PFSYKDLTSLGQSTKHFTIPTYLSFKHKNHIELDRL 235
 ++ G ER+N+L + + + F +K L SG STK FIAP+ L F + + +
 Sbjct: 191 VMDGWERLNLHGVYHPDGEIFNFDWKWLAPSGSLTKDFTIAPSSLCFGNAKTFGMGGKYG 250
 Query: 236 QIVYVRDYGMELGDKFIRDLMQSDLEVMISLHAKGSTKSETMTKLRTTKTLMESQKIGEQ 295
 + +++ EL D + D + ++ V+++LH + +++ + ++ K T +++ KI EQ
 Sbjct: 251 AVSFLQILSPELSDMLADFLNTESGVLVNLHVQAIEQTKAIKTIKRKITDLDAMKIAEQ 310
 Query: 296 QKMARTGIYLEKVGHVLENNIDEAEALLQTMTQTGDKLFDTVFLIGVLADTEDQLKQSLD 355
 +K R+G ++ + L ++A+ LL + ++LF FL+ +ADT+ +L +
 Sbjct: 311 KKAVERSGYDMILPSDLATYGEDAKKLLTKLQTRNERLFQLTFLVLNVADTKQKLNNDFV 370
 Query: 356 IIKQVAGSNDMIIDNLTMYQEAFAFNSLLPFGKNYLEGVSRLSLLTSNIAVNAPWTSVDIHD 415
 VA ++ + L Y QE S LP G N ++ + RSL TS++AV P+ + ++
 Sbjct: 371 QAAGVAQKHNCPLVRLDYQQEQGLASSLPLGVNQIK-IQRSLTSSVAVFVFPVFTQELFQ 429
 Query: 416 KGGK-FYGINQISSNIISIDRGKLNTPSGLILGTSGAGKGMATKHEIISTKLKEADSDTE 474
 G +YGIN S N+I +DR + P+ L LGT G+GK M+ K EI+S L D +
 Sbjct: 430 GGAAMYGINAKSRNMIMLDRKQARCPNALKLGTGSGKMSCKSEIVSVFLTPD---D 486
 Query: 475 IIIIVDPENEYSIIGQAFGGESIDIAPDSTTFLNVLELS-DENMDEDPVKVSEFLLSWIG 533
 I I DPE EY + + G+ I ++P S F+N L+++ + + D++P+ +KS+F+LS+
 Sbjct: 487 IFISDPEAEYYPLVKRLHGVIRLSPTS KDFVNPLDINLNYSEDDNPLALKSDFVLSFCE 546
 Query: 534 KLLDRK--MDGREKSLIDRVTRLTYKHF-----DTPSLVEWVFVLSQQPBEAKDLAL 584
 ++ K ++ EK++IDR R+ Y+ + + P L + L Q EA +A
 Sbjct: 547 LVMGGKNGLEAIEKTVIDRAVRVIYRPYLADPRPENMPILSDLHKALLDQHVPEADRAVAQ 606
 Query: 585 DMELYVEGSLDIFSHRTNIKTDSHPLIYNVKKLGDELKQIALMVIFDQIWNRVVKNQKLG 644
 ++LYV GSL++F+HRTN+ + + +++K+LG +LK++ ++++ DQIW RV N+ G
 Sbjct: 607 ALDLYVSGSLNVFNHRTNVDIGNRLVSFDIKELGKQLKKGLMIVQDQIWGRVTANRSQG 666
 Query: 645 KKTWIYFDEMQLLLLDKYASDFFPKLWSRVKYGAIPTGITQNVETLLLDANGRRRIANS 704
 K TW + DE LLL ++ + + ++W R RK+G IPTG TQNV+ LL I+ NS
 Sbjct: 667 KATWYFADEFHLLKKEQTAAYSABEIKRFRKNGGIPTGATQNVKDLLSSPEIENILENS 726
 Query: 705 EFMILLKQAKSDREELVHMLGLSKELEKYLVPKAGAGLIKAGSTVVPFKNKIPQHTKLF 764
 +F+ LL QA DR+ L L LS E +KY+ N E G GL+ + V+PF N IP +T+L+
 Sbjct: 727 DFITLLNQASGDRKILAERLNLSTEQQKYIDNSEPGEGLLIFENVVLPFTNPPIPHNTQLY 786
 Query: 765 DIMST 769
 IM+T
 Sbjct: 787 KIMTT 791

SEQ ID 8926 (GBS75) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 11; MW 89.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 20 (lane 6; MW 114.7kDa).

GBS75-GST was purified as shown in Figure 197, lane 8.

GBS329 was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 8; MW 89kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 174 (lane 2; MW 114kDa).

GBS329-GST was purified as shown in Figure 220, lanes 9 & 10.

-2140-

Example 1891

A DNA sequence (GBSx1999) was identified in *S.agalactiae* <SEQ ID 5875> which encodes the amino acid sequence <SEQ ID 5876>. Analysis of this protein sequence reveals the following:

5 Possible site: 33
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2442(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 15 vaccines or diagnostics.

Example 1892

A DNA sequence (GBSx2000) was identified in *S.agalactiae* <SEQ ID 5877> which encodes the amino acid sequence <SEQ ID 5878>. This protein is predicted to be DNA-directed RNA polymerase ii largest subunit. Analysis of this protein sequence reveals the following:

20 Possible site: 21
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4393(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 vaccines or diagnostics.

Example 1893

A DNA sequence (GBSx2001) was identified in *S.agalactiae* <SEQ ID 5879> which encodes the amino acid sequence <SEQ ID 5880>. Analysis of this protein sequence reveals the following:

35 Possible site: 13
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.92	Transmembrane	256 - 272 (250 - 277)
INTEGRAL	Likelihood = -8.28	Transmembrane	216 - 232 (213 - 244)
INTEGRAL	Likelihood = -8.12	Transmembrane	151 - 167 (148 - 191)
40 INTEGRAL	Likelihood = -7.27	Transmembrane	57 - 73 (54 - 80)
INTEGRAL	Likelihood = -6.74	Transmembrane	93 - 109 (88 - 111)
INTEGRAL	Likelihood = -3.50	Transmembrane	172 - 188 (168 - 191)
INTEGRAL	Likelihood = -2.76	Transmembrane	113 - 129 (110 - 130)

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database.

-2141-

>GP:AAG38039 GB:AF295925 Orf23 [Streptococcus pneumoniae]
 Identities = 71/86 (82%), Positives = 83/86 (95%)

Query: 37 VKSLADFNPTVWSYMTAITKGIMQPLGVAILAVVLVLEFSKMAKKIANS CGAMTFEAIAP 96
 +KSL+ +NPTVW+YM++ITK +MQPLGVAIL+VVL+LEFSKMAKKIANS CGAMTFEA+AP
 Sbjct: 1 MKSLSSYNPTVWTYMSITKSVMQPLGVAILSVVLILEFSKMAKKIANS CGAMTFEALAP 60

Query: 97 MIVSYIMVAVVITNTTVIVEAIIAIA 122
 M++SYIMVAVVITNTTVIVEAII IA
 Sbjct: 61 MLISYIMVAVVITNTTVIVEAIIIGIA 86

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

15 Example 1894

A DNA sequence (GBSx2002) was identified in *S.agalactiae* <SEQ ID 5881> which encodes the amino acid sequence <SEQ ID 5882>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

20	INTEGRAL	Likelihood = -7.54	Transmembrane	32 - 48 (25 - 52)
	INTEGRAL	Likelihood = -4.09	Transmembrane	63 - 79 (62 - 80)

----- Final Results -----

25	bacterial membrane	--- Certainty=0.4015(Affirmative) < succ>
	bacterial outside	--- Certainty=0.0000(Not Clear) < succ>
	bacterial cytoplasm	--- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9933> which encodes amino acid sequence <SEQ ID 9934> was also identified. A related GBS nucleic acid sequence <SEQ ID 10777> which encodes amino acid sequence <SEQ ID 10778> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 Example 1895

A DNA sequence (GBSx2003) was identified in *S.agalactiae* <SEQ ID 5883> which encodes the amino acid sequence <SEQ ID 5884>. This protein is predicted to be TrsK-like protein (traK). Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have a cleavable N-term signal seq.

40	INTEGRAL	Likelihood = -7.38	Transmembrane	66 - 82 (62 - 85)
----	----------	--------------------	---------------	--------------------

----- Final Results -----

45	bacterial membrane	--- Certainty=0.3951(Affirmative) < succ>
	bacterial outside	--- Certainty=0.0000(Not Clear) < succ>
	bacterial cytoplasm	--- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG38037 GB:AF295925 Orf21 [Streptococcus pneumoniae]
 Identities = 343/457 (75%), Positives = 385/457 (84%), Gaps = 24/457 (5%)

-2142-

Query: 142 LIVIGSGGAGKTRFRVKNLIQLNCSNIVVDPKDHLEAKTGKLFLENGYQVKVLDLVNMT 201
 + VIGGSG+GKTRFRVKNLIQ+N SNIVVDPKDHLEAKTGKLFLE+GYQVKVLDLVNM
 Sbjct: 1 MAVIGSGSGKTRFRVKNLIQMNSSNIVVDPKDHLEAKTGKLFLEHGYQVKVLDLVNMT 60

Query: 202 NSDGFNPFPRYVETENDLNRLMTVYFNNTKNGSGSRSDPFWDEASMTLVRAIASYLVDFYNP 261
 NSDGFNPFPRY+ETENDLNRLM VYFNNTKG+GSRSDPFWDEASMTLVRA+ASYLVDFYNP
 Sbjct: 61 NSDGFNPFPRYIETENDLNRLMAVYFNNTKSGSGSRSDPFWDEASMTLVRAIASYLVDFYNP 120

Query: 262 PGS-----SKQEQEARKRGRYPAFSEIGKLIKLLSKGDNQDKS 300
 P + K+E E R+KRGR F E + + + KS
 Sbjct: 121 PKTREQLIEESRLSQKEYQNLLKQKKEVEERKKRGRLSKFCESQNSLNTYPRVKTR-KS 179

Query: 301 ILEVLFEDYAKKYGHENFTMRNWADFQNYKDKTLDVIAVTTAKFALFNIQSVIDLTRD 360
 +LE+LFE+YAKKYG ENFTMRNWADFQNYKDKTLDVIAVTTAKFALFNIQSV+DLT+RD
 Sbjct: 180 VLEILFENYAKKYGTENFTMRNWADFQNYKDKTLDVIAVTTAKFALFNIQSVMDLTKRD 239

Query: 361 TMDLKTWGTQKTMVYLVIPDNDTFRFLSAL-FFSTVFSTLTQADV-DFKGQLPIHVR 418
 T+D+KTWG +K+MVYLVIPDND+TFRFLSAL FF+ F T + + + +LP+HVR
 Sbjct: 240 TLDKMTWQGEKSMVYLVIPDNDSTFRFLSALLFFNPYFQTPNKTSQILMLRVRLPLHVRV 299

Query: 419 YLDEFANVEIPDFAEQTSTVSRNMSLVPILQNIQLQGLYKEKEAWKTILGNCDLLY 478
 YLDEFAN+GEIPDFAEQTSTVSRNMSLVPILQNIQLQGLYKEKEAWKTILGNCDL+Y
 Sbjct: 300 YLDEFANIGEIPDFAEQTSTVSRNMSLVPILQNIQLQGLYKEKEAWKTILGNCDLVY 359

Query: 479 LGGNDEETFKFMSGLLGKQTVDVRSTSRSGQTGSSSTSHQKIARDLMTADEVGTMKRDE 538
 LGGNDE+TFKFMSGLLGKQT+DVR+TSRSGQTGS S SHQKIARDLMT DEVG MKR E
 Sbjct: 360 LGGNDEDTFKFMSGLLGKQTIDVRNTSRSGQTGSGSLSHQKIARDLMTPEVGNMKRHE 419

Query: 539 CLVRIAGVPVFRTKKYFPLKHKHKLADKETDDRWW 575
 CLVRIA +PVF++KKY KH +WK LA++ETD+R W
 Sbjct: 420 CLVRIANMPVFKSKKYNSTKHPNWKYLANQETDERRW 456

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8929> and protein <SEQ ID 8930> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 5
 McG: Discrim Score: 5.53
 GvH: Signal Score (-7.5): -0.78
 Possible site: 34
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 1 value: -7.38 threshold: 0.0
 INTEGRAL Likelihood = -7.38 Transmembrane 66 - 82 (62 - 85)
 PERIPHERAL Likelihood = 1.75 338
 modified ALOM score: 1.98

*** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

33.9/50.9% over 419aa

Lactococcus lactis

GP|3582206| trsK protein (traK) Insert characterized

PIR|T43089|T43089 transfer complex protein TrsK - plasmid pMRC01 Insert characterized

ORF00383(715 - 2004 of 2415)

-2143-

GP|3582206|gb|AAC56002.1|AE001272(23 - 442 of 530) trsK protein (traK) {Lactococcus
lactis}PIR|T43089|T43089 transfer complex prote
in TrsK - Lactococcus lactis plasmid pMRC01

%Match = 10.1

%Identity = 33.8 %Similarity = 50.8

Matches = 141 Mismatches = 193 Conservative Sub.s = 71

519 549 579 609 639 669 699 729
SFLA FILGLV LMM TLVY LVSTG QKVYREGE EYG SARFG TSKEK RNFYS KNPN FNDT IILARD VRLTL LEKKKPQ FDRNKNLI
| : | ::
MNGT ILGLVD NKLII YQDNT TTKPNRNV M
10 20

759 789 816 846 876 906 936 966
VIGGSGAGKTRFRVKPNLIQLNCSNIVV-DPKDHLAEKTKGLFLENGYQVKVLDLVNMTNSDGFNPFRRVETENDLNRL
|||||: || || :||| ||| ||| :| ||:| ::| || :||| :| : :
VIGGSSGYKTQSVVITNLFNETKNSIVVTDPKGELYEKTAGIKLAQGYEVHVHVFANMAHSDRYNPFDYIERDIQAESVA
40 50 60 70 80 90 100

996 1026 1056 1086 1116 1146 1176 1194
TVYFNNITKGNGSRSDPFWDEASMTLVRAIASYLVDFYNPPGSSKQEQEARRKGRYPAFSEIGKLIKLLSKGD----NQD
| : | : | | :: | : :: : | | : : | : | | | : |
TKIVQSENAEGKK--DVWFSTQRQLLKALILFVM-----KERSPEQRNLGAVINVLTQTFDSEPIKND
 120 130 140 150 160

[illegible]

1461 1491 1521 1551 1581 1611 1641 1671
IPDNDTTRFRFLSALFFSTVFSSTLTRQADVDFKGLPIHVRSYLDEFANVEGIPDFAEQTSTVRSRNMSLVPILONIAQLQ
|| | |: : ||| : | : : || | | |: : | : : | : : | : ||
IPVMDNTYESFINLFFSQMFDELYKLASSN-GAKLPQEVDFILDFEVNLGKFPKYEEFLATCRGYGIGVITTCQTLTLOLQ

260 270 280 290 300 310 320

1701 1731 1761 1791 1809 1839 1869 1899
GLYKEKEAWKTLGNCDLLYLGNDEETFKFMSGLLGKQKIVDVR---STSRSGQTGSSSTSHQKIARDLMTADEVGT
|| || :||| : | :| :| ||| || | ||| | : | | : | ||| ||:
SLY-GKEKAESILGNHAVKICLNASNEATAKYFSELLGKSTVKVETGSESTSHSKETSTSKSDSYSYTSRQLMTPDEIIR

340 350 360 370 380 390 400

1929 1956 1974 2004 2034 2064 2094 2124
MKRDECLVRIAGV-PVFRTK---KYFPLKHKKWKLADKETDDRWNNYHINPLAKEEBELDSYQIRDLSTETSLH**K
| : | : | : || | | | : | : : : :
MPDTSQLLIFTNQPIKAATKAFQKLFPPDADSKVKLEQNKYVGITKSQLEKYNDLSVKWEELQSCLKNITVTVEEEKDL

420 430 440 450 460 470 480

SEQ ID 5884 (GBS11d) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 151 (lane 6; MW 61kDa) and in Figure 182 (lane 10; MW 61kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 12 (lane 5; MW 91.5kDa).

Example 1896

A DNA sequence (GBSx2004) was identified in *S.agalactiae* <SEQ ID 5885> which encodes the amino acid sequence <SEQ ID 5886>. Analysis of this protein sequence reveals the following:

Possible site: 50

```
>>> Seems to have no N-terminal signal sequence
```

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4192 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

```

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

-2144-

A related GBS nucleic acid sequence <SEQ ID 9935> which encodes amino acid sequence <SEQ ID 9936> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1897

10 A DNA sequence (GBSx2005) was identified in *S.agalactiae* <SEQ ID 5887> which encodes the amino acid sequence <SEQ ID 5888>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
15      bacterial cytoplasm --- Certainty=0.3391(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1898

25 A DNA sequence (GBSx2006) was identified in *S.agalactiae* <SEQ ID 5889> which encodes the amino acid sequence <SEQ ID 5890>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have an uncleavable N-term signal seq

30      INTEGRAL    Likelihood = -10.03    Transmembrane    68 - 84 ( 64 - 90)
      INTEGRAL    Likelihood = -7.06     Transmembrane    33 - 49 ( 27 - 50)
      INTEGRAL    Likelihood = -5.73     Transmembrane   106 - 122 ( 105 - 123)
      INTEGRAL    Likelihood = -4.46     Transmembrane     6 - 22 ( 3 - 24)
      INTEGRAL    Likelihood = -2.13     Transmembrane   154 - 170 ( 154 - 170)
      INTEGRAL    Likelihood = -0.53     Transmembrane   180 - 196 ( 180 - 196)

35      ----- Final Results -----
      bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

40 A related GBS nucleic acid sequence <SEQ ID 9937> which encodes amino acid sequence <SEQ ID 9938> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA11325 GB:D78257 ORF8 [Enterococcus faecalis]
Identities = 35/102 (34%), Positives = 57/102 (55%), Gaps = 4/102 (3%)

45      Query: 90   TRNQAVLVQVGKQVPPPIIFLLFL-VNASILEEIVYRQLLWEKLTFF--PFEQIGVTSFLFV 146
      T N + L+++   V P++ +L L + A I+EEIV+R + L I ++SFLF
      Sbjct: 7    TANDSTLIKLFSGVSPVLVLLLGIAAPIMEEIVFRGGIIGYLVENNALLAILISSFLFG 66
```

Query: 147 LSHGPNQLGSWLIYSCLGLTLAVVRLKT-DCMTAIALHLLWN 187
+ HGP S+ +Y +G+ L+V KT D +I++H L N
Subject: 67 IIHGPTNFISFGMYFFMGIILSVSYKTKDLRSISIHFLNN 108

A related GBS gene <SEQ ID 8931> and protein <SEQ ID 8932> were also identified. Analysis of this protein sequence reveals the following:

```

bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

774 804 831 861 891 921 951 981
HGP NQLGSLWIYISCLGLTAVRLKT-DCMTAIALHLWNLSLAYVVTFL*YQNQECFRIMEAPYV**GIEKRGGHYVI*T
||| : |: : | ::| :| || | ::::| |
HGPTNFISFGMYFFMGII LSVSYKYTKDLRVSI SIHF LNNLFPAIAIAYGLI
80 90 100 110 120

Possible site: 23

-2146-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2490(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9939> which encodes amino acid sequence <SEQ ID 9940> was also identified.

10 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1900

15 A DNA sequence (GBSx2008) was identified in *S.agalactiae* <SEQ ID 5893> which encodes the amino acid sequence <SEQ ID 5894>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.5298(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC98423 GB:L29323 unknown [Streptococcus pneumoniae]
 Identities = 68/126 (53%), Positives = 88/126 (68%)

30 Query: 1 MNLHKKSI LDC TELEER IHQAET NQL LQKILSLN FDCDFE VTFEDDYHKEMNDPLFYE 60
 M L+K+SI LDC ELE +H AE QL ++I +PN+ C+FEVTF DDYHK+ N PLFYE
 Sbjct: 1 MKALNKESILDCDELETELHDAEIKQLDEQIFLMPNYPCEFEVTFLLDDYHKHKNYPLFYE 60

 Query: 61 SNLHQISDFMETRDIKNGVDTLTKDNHLAFRAFGENYSARGKEGILTTLVTVKCFGEGR 120
 S L I +F+E++DIKNG D + +L F +G+ Y A GKEGILTT VTVK F E +
 35 Sbjct: 61 SYLQNI MEFL ESQDIKNGADAFVDDHQNLFVFLYQGQYRAEGKEGILTTQVTVKAFDEDK 120

Query: 121 MPIDMS 126

PI+ +

Sbjct: 121 KPINF A 126

40

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1901

45 A DNA sequence (GBSx2009) was identified in *S.agalactiae* <SEQ ID 5895> which encodes the amino acid sequence <SEQ ID 5896>. This protein is predicted to be methyl transferase. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

50

----- Final Results -----

-2147-

bacterial cytoplasm --- Certainty=0.1209(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC98421 GB:L29323 methyl transferase [Streptococcus pneumoniae]

Identities = 323/449 (71%), Positives = 389/449 (85%), Gaps = 3/449 (0%)

10 Query: 1 MKFLDLFAGIGGFRLGMESQGHKCLGFCEIDKFARTSYKAMFNTEGEIEYHDIKEVTDHD 60
 M+F+DLF+GIGGFRLGMES GH+C+GFCEIDKFAR SYK++F TEGEIE+HDI++V+D +
 Sbjct: 1 MRFIDLFSGIGGFRLGMESVGHCEIGFCEIDKFARESYKSIFQTEGEIEFHDIRDVSDD 60

15 Query: 61 FRQFRGQVDIICGGFPCQAFSLAGRRLLGFEDTRGTLFFEIARAQKIQPRFLFLENVKG 120
 F++ RG+VD+ICGGFPCQAFS+AGRRLLGFEDTRGTLFFEIARAQKIQPRFLFLENVKG
 Sbjct: 61 FKKLRGKVDVICGGFPCQAFSIAGRRLLGFEDTRGTLFFEIARAQKIQPRFLFLENVKG 120

20 Query: 121 LNHDEGRFTATILSTLDELGYDVEWQVLNSKDFQVPQNRERVFIIGHSRYSRPFIFPLR 180
 LNH+GRFT TIL+TLDELG+DVEWQ+LNSKDF VPQNRERVFIIGHSR+ +R FP R
 Sbjct: 121 LNHDKGRFTTILSTLDELGFDVWQMLNSKDFQVPQNRERVFIIGHSRKRGTRLGFPFR 180

25 Query: 181 RED---SPAHLERLGNINPSKHLNGEVYLTSLAPTLTRGKGEGAKIAIPVLTDPDRLEK 237
 RE +P L+ LGN+NPSK G++G+VY + GLAPTL RGKGEK KIAIP +TPDRL+K
 Sbjct: 181 REGQATNPETLKLGNLNPSSGMSGKVYYSEGLAPTLVRGKGEGFKIAIPCMTPDRLDK 240

30 Query: 238 RQHGRRFKDNQDPMFTLTSQDKHGVVAGNLPTSFDQTRGVFDISGLSPTLTTMQGGDKV 297
 RQ+GRRFKDNQ+PMFTL +QD+HG+VV G+LPTSF +TGRV+ GLSPTLTTMQGGDK+
 Sbjct: 241 RQNGRRFKDNQEPMTLNTQDRHGIVVVGDLPTSFKETGRVYGSEGLSPTLTTMQGGDKI 300

35 Query: 298 PKILLREELPFLKIKEATKTGYAKATLGDSVNLAYPDSTKRRGRVKGISNTLTSDNMG 357
 PKIL+ E + FLK++EATK GYA+A +GDS+NL P S RRGVKGKI+NTLTTS MG
 Sbjct: 301 PKILPEPIQFLKVREATKKGYAQAEIGDSINLERPSSQHRGRVKGIANLTLTSGQMG 360

40 Query: 358 VVVAALEYRQDKWYEVTVGIVLEGLYRLRIRRLTPRECFLRQGFDPWAYERAESVSSKSQ 417
 VVVA+ E + Y+V G++++G+ YRLRIR+TP+ECFLRQGFDPWA+E A VSS SQ
 Sbjct: 361 VVVASYEGEDKQVYQVAGVLIDGQFYRLRIRRLTPKECFRLQGFDPWAFEAARKVSSNSQ 420

45 Query: 418 LYKQAGNSVTVTVIEAIAREFRTEEEK 446
 LYKQAGNSVTV VI AIA++ + EE+++
 Sbjct: 421 LYKQAGNSVTVPVIAIAAKLKEVEEKDE 449

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2435> which encodes the amino acid sequence <SEQ ID 2436>. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1725(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 60/75 (80%), Positives = 69/75 (92%)

55 Query: 1 MKFLDLFAGIGGFRLGMESQGHKCLGFCEIDKFARTSYKAMFNTEGEIEYHDIKEVTDHD 60
 MKFLDLFAGIGGFRLG+ +Q H+C+GFCEIDKFAR SYKA++ TEGEIE+HDI++VTD D
 Sbjct: 4 MKFLDLFAGIGGFRLGLINQCHECIGFCEIDKFARQSYKAIYETEGEIEFHDIRQVTDQD 63

60 Query: 61 FRQFRGQVDIICGGF 75
 FRQ RGQVDIICGGF
 Sbjct: 64 FRQLRGQVDIICGGF 78

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2148-

Example 1902

A DNA sequence (GBSx2010) was identified in *S.agalactiae* <SEQ ID 5897> which encodes the amino acid sequence <SEQ ID 5898>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 16
   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -9.71    Transmembrane    8 - 24 ( 3 - 30)

      ----- Final Results -----
10      bacterial membrane --- Certainty=0.4885(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9941> which encodes amino acid sequence <SEQ ID 9942> was also identified.

15 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5899> which encodes the amino acid sequence <SEQ ID 5900>. Analysis of this protein sequence reveals the following:

```

20   Possible site: 28
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.81    Transmembrane    20 - 36 ( 19 - 36)

      ----- Final Results -----
25      bacterial membrane --- Certainty=0.1723(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 16/33 (48%), Positives = 26/33 (78%)

30   Query: 1  MNKMIWWILGGIYLLISIIILIVEIIRAPEMDDH 33
      ++KM WW+L G++ + I LI+E+I APEM+D+
      Sbjct: 12 VSKMFWWLLLGWGLRTIWLIIIEVITAPEMEDY 44

```

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1903

A DNA sequence (GBSx2011) was identified in *S.agalactiae* <SEQ ID 5901> which encodes the amino acid sequence <SEQ ID 5902>. This protein is predicted to be ifn-response binding factor 1 (irebf-1). Analysis of this protein sequence reveals the following:

```

40   Possible site: 53
   >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
45      bacterial cytoplasm --- Certainty=0.4771(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

50   >GP:AAD41248 GB:AF106927 unknown [Streptococcus suis]
      Identities = 258/272 (94%), Positives = 266/272 (96%)

      Query: 1  MKRITANQYQTSERYYKLPKILFESERYKDMKLEVKVAYAVLKDRLELSLSKGWIDEDGA 60
      MKRITANQYQTSERYYKLPKILFESERYKDMKLEVKVAYAVLKDRLELSLSKGWIDEDGA

```

-2149-

Sbjct: 1 MKRITANQYQTSERYKLPKILFESERYKDMKLEVKVAVLKDRLLESLSKGWIDEDGA 60

Query: 61 IYLIYSNSNLMALLGCSSKLLSIKKTIREYGLIDEVQSSSERGRMANKIYLGELEHEP 120
IYLIYSNSNLMALLGCSSKLLSIKKTIREYGLIDEVQSSSE+GRMANKIYLGELEHE

5 Sbjct: 61 IYLIYSNSNLMALLGCSSKLLSIKKTIREYGLIDEVQSSSEKGRMANKIYLGELEHET 120

Query: 121 TPVLHTDGASVKKTLGESQKRTGPNLYSAPSETEGSETKYSETEGSDLVMDDEERQLVD 180
TPVLHTDGASVKKTLG SQKRTGPNL SAPSETEGSETKYSET+GSD +++DEEERQ VD

10 Sbjct: 121 TPVLHTDGASVKKTLGGSQKRTGPNLSAPSETEGSETKYSETKGSDFLIEDEEERQQVD 180

Query: 181 EKKEENFTSKVDGVTKYDRDYIWGLVHDQLRQTGLSQSASDYAMIYFSDRYQYALEQMRF 240
EK+EENFTSKVDGVT+YDRDYIWGLVHDQLRQTGLSQSASDYAMIYFSDRYQYALE MRF

Sbjct: 181 EKQEENFTSKVDGVTRYDRDYIWGLVHDQLRQTGLSQSASDYAMIYFSDRYQYALEHMR 240

15 Query: 241 ARSAEVIAEYVFNGVLSEWTKQLRRQEVKGG 272
ARSAEVIAEYVFNGVLSEWTKQLRRQEVKGG+

Sbjct: 241 ARSAEVIAEYVFNGVLSEWTKQLRRQEVKGGD 272

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5903> which encodes the amino acid
20 sequence <SEQ ID 5904>. Analysis of this protein sequence reveals the following:

Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.5248(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 84/122 (68%), Positives = 99/122 (80%), Gaps = 2/122 (1%)

Query: 145 VLYSAPSETEGSETKYSETEGSDLVMDDEERQLVD--EKKEENFTSKVDGVTKYDRDYI 202
VL SAPSETE SET+ SET+ S+LV++DEEER+ +K E +FT +VD VTKYD+DYI

35 Sbjct: 1 VLNSAPSETEKSETEGSETKESNLVIEDEEERKECTSVKKTEGHFTRQVDQVTKYDKDYI 60

Query: 203 WGLVHDQLRQTGLSQSASDYAMIYFSDRYQYALEQMRFARSAEVIAEYVFNGVLSEWTKQ 262
W LVH QLR+ GLSQ+ASD M YF +RY YALE +RFAR+AE IAEYVFNGVLSEWTKQ

Sbjct: 61 WSLVHSQLEGGLSQAASDLVMSYFEERYAYALEHIRFARTAEATAEYVFNGVLSEWTKQ 120

40 Query: 263 LR 264
LR

Sbjct: 121 LR 122

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
45 vaccines or diagnostics.

Example 1904

A DNA sequence (GBSx2012) was identified in *S.agalactiae* <SEQ ID 5905> which encodes the amino
acid sequence <SEQ ID 5906>. Analysis of this protein sequence reveals the following:

Possible site: 17
50 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4191(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
55 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9375> which encodes amino acid sequence <SEQ ID 9376>
was also identified.

-2150-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 1905

A DNA sequence (GBSx2013) was identified in *S.agalactiae* <SEQ ID 5907> which encodes the amino acid sequence <SEQ ID 5908>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3723(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 1906

A DNA sequence (GBSx2014) was identified in *S.agalactiae* <SEQ ID 5909> which encodes the amino acid sequence <SEQ ID 5910>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3053(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 Example 1907

A DNA sequence (GBSx2015) was identified in *S.agalactiae* <SEQ ID 5911> which encodes the amino acid sequence <SEQ ID 5912>. This protein is predicted to be 50S ribosomal protein L7/l12 (rplL). Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1034(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```


-2151-

A related GBS nucleic acid sequence <SEQ ID 9943> which encodes amino acid sequence <SEQ ID 9944> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:CAB11881 GB:Z99104 ribosomal protein L12 (BL9) [Bacillus subtilis]
Identities = 83/123 (67%), Positives = 95/123 (76%), Gaps = 2/123 (1%)

Query: 6 MALNIENIIAEIKEATILELNDLVKAIIEEFVGTAAAPVAAA--AAGGEAAAAKDSFDVE 63
MALNIE IIA +KEAT+LELNDLVKAIIEEFVGTAAAPVA A AA G AA + FD+

10 Sbjct: 1 MALNIEEIIASVKEATVLELNDLVKAIIEEFVGTAAAPVAVAGGAAAGGAAEEQSEFDLI 60

Query: 64 LTAAGDKKVGVIKVVREITGEGLEAKAIVDNAPSVIKEGASEAEANEIKEKLEAAGASV 123
L AG +K+ VIKVVREITG GLKEAK +VDN P +KEG ++ EA E+K KLE GASV

15 Sbjct: 61 LAGAGSQIKIKVIKVVREITGLGLKEAKELVDNTPKPLKEGIAKEEAEELKAKLEEVGASV 120

Query: 124 TLK 126
+K

Sbjct: 121 EVK 123

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5913> which encodes the amino acid sequence <SEQ ID 5914>. Analysis of this protein sequence reveals the following:

Possible site: 51
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1164(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below.

Identities = 104/126 (82%), Positives = 113/126 (89%)

Query: 1 MEEITMALNIENIIAEIKEATILELNDLVKAIIEEFVGTAAAPVAAAAAGGEAAAADSF 60
+EEITMALNIENIIAEIKEA+ILELNDLVKAIIEEFVGTAAAPVAAAAAGG AAKDSF

35 Sbjct: 1 LEEITMALNIENIIAEIKEASILELNDLVKAIIEEFVGTAAAPVAAAAAGGAEAAKDSF 60

Query: 61 DVELTAAGDKKVGVIKVVREITGEGLEAKAIVDNAPSVIKEGASEAEANEIKEKLEAAG 120
DVELT+AGDKKVGVIK VREITG GLKEAK +VD AP+ +KEG + AEA EIK KLE AG

40 Sbjct: 61 DVELTSAGDKKVGVIKAVREITGLGLKEAKGLVDGAPNVKEGVAAAEAEIKAKLEEAG 120

Query: 121 ASVTLK 126
A++TLK

Sbjct: 121 ATTTLK 126

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1908

A DNA sequence (GBSx2017) was identified in *S.agalactiae* <SEQ ID 5915> which encodes the amino acid sequence <SEQ ID 5916>. This protein is predicted to be ribosomal protein L10 (rplJ). Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1251(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-2152-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:CAB11880 GB:Z99104 ribosomal protein L10 (BL5) [Bacillus subtilis]
Identities = 96/164 (58%), Positives = 125/164 (75%), Gaps = 1/164 (0%)

Query: 14 MSEAIIAKAEQVELIAEKMKAASIVVDSRGLTVEQDTNLRRLRESDEVFKVIKNSI 73
MS AI KK VE IA K+K + S ++VD RGL V + T LR+ LRE++VE KV KN++
10 Sbjct: 1 MSSAIETKKVV-VEEIASKLKESKSTIIVDYRGLNVSEVTELKQLREANVESKVYKNTM 59

Query: 74 LTRAAEKAGLEDLKELFVGPSAVAFSNEDVIAPAKVISDFAKDAEAEIKGGSVDGKFTS 133
RA E+A L L + GP+A+AFS EDV+APAKV++DFAK+ EAEIK G ++GK ++
Sbjct: 60 TRRAVEQAEINGLNDFLTGPNAIAFSTEDVAPAKVLNDFAKNHEAEIKAGVIEGKVST 119

15 Query: 134 VEEINALAKLPNKEGMLSMLLSVLQAPVRNVAYAVKAVAEKDEE 177
VEE+ ALA+LP +EG+LSMLLSVL+APVRN+A A KAVAE+ EE
Sbjct: 120 VEEVKALAELEPPREGLLSMLLSVLKAPVRNLALAAKAVAEQKEE 163

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5917> which encodes the amino acid sequence <SEQ ID 5918>. Analysis of this protein sequence reveals the following:

Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -5.47 Transmembrane 7 - 23 (5 - 24)

25 ----- Final Results -----
bacterial membrane --- Certainty=0.3187(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below.

Identities = 149/176 (84%), Positives = 162/176 (91%)

Query: 4 SQKIKTEVKLMSEAIIAKAEQVELIAEKMKAASIVVDSRGLTVEQDTNLRRLRES 63
S KIKTEVKLMSEAIIAKAEQVELIAEKMKAASIV+VDSRGLTV+QDT LRRSLRES
35 Sbjct: 23 SPKIKTEVKLMSEAIIAKAEQVELIAEKMKAASIVIVDSRGLTVQDTVLRRSLRESG 82

Query: 64 VEFKVIKNSILTRAAEKAGLEDLKELFVGPSAVAFSNEDVIAPAKVISDFAKDAEAEIK 123
VEFKVIKNSILTRAAEKAGL++LK++FVGPSAVAFSNEDVIAPAKVI+DF K A+ALEIK
40 Sbjct: 83 VEFKVIKNSILTRAAEKAGLEDLKDVFVGPSAVAFSNEDVIAPAKVINDFTKTADALEIK 142

Query: 124 GGSVDGKFTSVEEINALAKLPNKEGMLSMLLSVLQAPVRNVAYAVKAVAEKDEEVA 179
GG+++G +S EEI ALA LPN+EGMLSMLLSVLQAPVRNVAYAVKAVAE E A
Sbjct: 143 GGAIEGAVSSKEETQALATLPNREGMLSMLLSVLQAPVRNVAYAVKAVAEKDEGAA 198

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1909

A DNA sequence (GBSx2018) was identified in *S.galactiae* <SEQ ID 5919> which encodes the amino acid sequence <SEQ ID 5920>. Analysis of this protein sequence reveals the following:

50 Possible site: 40
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -7.22 Transmembrane 125 - 141 (106 - 143)
INTEGRAL Likelihood = -1.91 Transmembrane 108 - 124 (106 - 124)

55 ----- Final Results -----
bacterial membrane --- Certainty=0.3888(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-2153-

A related GBS nucleic acid sequence <SEQ ID 10931> which encodes amino acid sequence <SEQ ID 10932> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1910

- 10 A DNA sequence (GBSx2019) was identified in *S.agalactiae* <SEQ ID 5921> which encodes the amino acid sequence <SEQ ID 5922>. This protein is predicted to be Clp-like ATP-dependent protease binding subunit (clpC). Analysis of this protein sequence reveals the following:

Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 15 bacterial cytoplasm --- Certainty=0.3483(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 20 >GP:AAA68910 GB:L34677 Clp-like ATP-dependent protease binding
subunit [Bos taurus]
Identities = 437/589 (74%), Positives = 514/589 (87%), Gaps = 5/589 (0%)
- 25 Query: 10 DPFNG-MDDIFNSLMGNMGVNSENKRYLINGREVTPEEFSQYRQTGKLPQQLNNQNT 68
DPF N MDD+FN LMG M G NSEN+RYLINGREVTPEE++ +RQTGKLPG Q
Sbjct: 2 DPFNNDMDLFLNQLMGGMNGVNSENRRYLINGREVTPEEYAFRQTGKLPGVTDPTQ-AK 60
- 30 Query: 69 TNQVSADSVLTKLGTNLTDQARQHLLDLPVIGRNKEIQETAELARRTKNNPVLVG DAGVG 128
T Q DS+L KLG NLT +A++ LDPVIGRNKEIQETAEL+RRTKNNPVLVG DAGVG
Sbjct: 61 TKQPQDSMLAKLGRNLTQEAKEGKLDLPVIGRNKEIQETAELSRRTKNNPVLVG DAGVG 120
- 35 Query: 129 KTAVIEGLAQAIINGDVPAAIKNEIISIDISSLEAGTQYRGSFEENIQNIIEVKETGN 188
KTAV+EGLAQAI+ GDVPAAIKKN+IISIDISSLEAGTQYRGSFEEN+Q +I EVK+ GN
Sbjct: 121 KTAVVEGLAQAI VAGDVPAAIKKNQIISIDISSLEAGTQYRGSFEENMQKLIDEVKKDGN 180
- 40 Query: 189 IILFFDEIHQILGAGSTGGDSGSKGLADILKPAISRGELTVIGATTQDEYRNTILKNAAL 248
+ILFFDEIHQI+GAG+ G SGSKG+ADILKPAISRGELT+IGATTQDEYRNTILK+AAAL
Sbjct: 181 VILFFDEIHQIIGAGNAGDASGSKGMADILKPAISRGELTVIGATTQDEYRNTILKDAAL 240
- 45 Query: 249 ARRFNEVKVNAPSQDTFNILMGIRNLYEQHHNVVLPDSVLKAAVDLSIQYIPQRSPLDK 308
+RRFN+V VNAPS +DTF IL G+R LYE+HHNV LPD VLKAA+D S+QYIPQRSPLDK
Sbjct: 241 SRRFNQVTVNAPSKEDTFKILQGLRKLKYEKHHNVSLPDEVLKAADYSVQYIPQRSPLDK 300
- 50 Query: 309 AIDLIDMTAAHLAAQHPVTDLKSLEKEIAEQRDQKEKAVNTEDFERALKVKTRIEELQ 368
AIDLID+TAAHLA++HPV D K++E+EI + KQ++AV ED++ A + K ++ +LQ+Q
Sbjct: 301 AIDLIDVTAHLASKHPVKDAKTIEEIKKTEAKQEQEAVEKEEDYQAAQEAQDQVAKLQDQ 360
- 55 Query: 369 IDNHTGQKVTTATINDIAMSIERLTGVPVSNMGASDIERLKGKLVIGQND AVEA 428
+ +H+E ++V AT +D+A. ++ER+TG+PVS MGASDIERLK L RL+GKLVIGQ +AVEA
Sbjct: 361 LKDHSESERVVATPSDVAAAVERTGIPVSKMGASDIERLKGKLVIGQND AVEA 420
- Query: 429 VARAIRNRNAGFDDGNRPIGSFLFVGPTGVGKTELAKQLAFDMFGSKDAIVRLDMSEYND 488
V+RAIRNRNAGFD+GNRPIGSFLFVGPTGVGKTELAKQLA DMFGS + I+RLDMSEY D
Sbjct: 421 VRAIRNRNAGFDEGNRPIGSFLFVGPTGVGKTELAKQLALDMFGSTNDIIRLDMSEYTD 480
- Query: 489 RTAVSKLIGATAGYVGYDDNSNTLTERIRRNYPYSIVLLDEIEKADPQVITLLQLVLD DGR 548
RTAVSKLIG TAGYVGYDDNSNTLTE++RR+PYSIVLLDEIEKA+PQVITLLQLVLD DGR

-2154-

Sbjct: 481 RTAVSKLIGTTAGYVGYYDDNSNTLTEKVRHPYSIVLLDEIEKANPQVITLLQLVDDGR 540

Query: 549 LTDGGQNTINFKNVTIIATSNAGFGNEAFTGDSKDLKIMERISPYFRP 597

LTDGGQNT++FKNT+IIATSNAGF ++A G+ D K+M+++ PYFRP

5 Sbjct: 541 LTDGGQNTVDFKNTIIATSNAGFSSDAVAGE---DAKLMDKLQPYFRP 586

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5923> which encodes the amino acid sequence <SEQ ID 5924>. Analysis of this protein sequence reveals the following:

Possible site: 22

10 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2718(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

15 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 551/697 (79%), Positives = 616/697 (88%), Gaps = 3/697 (0%)

20 Query: 5 NFYNRDPFGNMDDIFNLSMGMGGYSENKRYLINGREVTPEEFSQYRQTGKLPQQLNN 64
+F +DPF NMDDIFN LM NMGGY SEN RYL+NGRE+TPEEF YRQTG+LP
Sbjct: 3 HFSGKDPFVNMMDDIFNQLMANMGGYRSENPRYLNGREITPEEFQHYRQTGQLPVATTKA 62

25 Query: 65 QNTPTNQVSADSVLTKLGTNLTDQARQHLLDPVIGRNKEIQETAELARRTKNNPVLVGD 124
N+ ADSVLT+LGTNLT +ARQ LDPVIGRNKEIQ+TAEILARRTKNNPVLVGD
Sbjct: 63 TNSQMLTPKADSVLTQLGTNLTDQARQHLLDPVIGRNKEIQDTAEILARRTKNNPVLVGD 122

30 Query: 125 AGVGKTAVIEGLAQAIINGDVPAAIKNKEIISIDISSLEAGTQYRGSFEENIQNIIEVK 184
AGVGKTAVIEGLAQAI+NGDVPAAIKNKEI+SIDISSLEAGTQYRGSFEE IQN+I+EVK
Sbjct: 123 AGVGKTAVIEGLAQAI+NGDVPAAIKNKEI+SIDISSLEAGTQYRGSFEETIQNLIQEVK 182

35 Query: 185 ETGNIILFFDEIHQILGAGSTGGSGSKGLADILKPALSRGELTVIGATTQDEYRNTILK 244
E GNIILFFDEIHQI+GAG+T DSGSKGLADILKPALSRGELT+IGATTQDEYRNTILK
Sbjct: 183 EAGNIILFFDEIHQIVGAGATSSDSGSKGLADILKPALSRGELTLIGATTQDEYRNTILK 242

40 Query: 245 NAALARRFNEVKVNAPSAQDTFNILMGIRNLYEQHHNVLPDSVLKAAVDLSIQYIPQRS 304
NAALARRFNEVKVNAPSA+DTF+ILMGIRNLYEQHH++ LPD+VLKAAVD SIQYIPQRS
Sbjct: 243 NAALARRFNEVKVNAPSAEDTFHILMGIRNLYEQHHITLPDNLKAAVDYSIQYIPQRS 302

45 Query: 305 LPDKAIDLIDMTAAHLAAQHPVTDLKSLEKETAEQRDKQEKAVNTEDFEEALKVKTRIEE 364
LPDKAIDL+DMTAAHLAAQHPVTDLK+LE EIA+Q++ QEKAV EDFE+AL KTRIE
Sbjct: 303 LPDKAIDLIDMTAAHLAAQHPVTDLKTLETETAKQKESQEKAVAKEDFEKALAAKTRIE 362

50 Query: 365 LQNQIDNHTEGQKVATINDIAMSIERLTGVPVSNMGASDIERLKGELGNRLKGKGVIGQND 424
LQ QI+ H + Q VTAT+NDIA S+ERLTG+PVSNMG +D+ERLK + +RLK VIGQ++
Sbjct: 363 LQKQIEQHNSQNVATVNDIABSVRLTGIPVSNMGINDLERLKGISRLKSHVIGQDE 422

55 Query: 425 AVEAVARAIRNRAGFDDGNRPIGSFLFVGPTGVGKTELAKQLAFDMFGSKDAIVRLDMS 484
AV AVARAIRNRAGFDDG RPIGSFLFVGPTGVGKTELAKQLA D+FGSKDAI+RLDMS
Sbjct: 423 AVAVARAIRNRAGFDDGKRPIGSFLFVGPTGVGKTELAKQLALDLFGSKDAIIRLDMS 482

60 Query: 485 EYNDRTAVSKLIGATAGYVGYYDDNSNTLTERIRRNYPYSIVLLDEIEKADPQVITLLQLVL 544
EYNDRTAVSKLIG TAGYVGYYDDN+NTLTER+RRNPY+IVLLDEIEKADPQ+ITLLQLVL
Sbjct: 483 EYNDRTAVSKLIGTTAGYVGYYDDNNNTLTERVRRNPYAIVLLDEIEKADPQIITLLQLVL 542

65 Query: 545 DDGRLTDGQNTINFKNVTIIATSNAGFGNEAFTGDSKDLKIMERISPYFRPEFLNRFN 604
DDGRLTDGQNTINFKNVTIIATSNAGFG + + IM+RI+PYFRPEFLNRFN
Sbjct: 543 DDGRLTDGQNTINFKNVTIIATSNAGFGQQ---DTETSESINIMDRIAPYFRPEFLNRFN 599

Query: 605 GVIEFSLSKDDLSEIVDLMLDEVNQITGKKGIDLVVDENVKSHLIELGYDEAMGVRPLR 664
+I+F+HL K+ L EIVDLML EVNQIT KKEI L + ++ K+HLI+LGY+ AMG RPLR
Sbjct: 600 SIIKFNLQKESLEEIVDLMLAEVNQITAKKGISLITDDAKAHLIDLGYNHAMGARPLR 659

Query: 665 RVIEQEIRDRTIDYYLDHTDVKHLKANLQDQQIVISE 701
R+IEQEIRDRTIDYYLDH +VK L+A L++GQ+VI +

-2155-

Sbjct: 660 RIIEQEIRDITDYLDHPEVKKLQAILKEGQLVIRQ 696

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 1911

A DNA sequence (GBSx2020) was identified in *S.agalactiae* <SEQ ID 5925> which encodes the amino acid sequence <SEQ ID 5926>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have an uncleavable N-term signal seq
 10 INTEGRAL Likelihood = -4.78 Transmembrane 8 - 24 (7 - 25)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2911(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 15 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9945> which encodes amino acid sequence <SEQ ID 9946> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AAC73364 GB:AE000134 putative enzyme [Escherichia coli K12]
 Identities = 142/307 (46%), Positives = 195/307 (63%), Gaps = 6/307 (1%)
 Query: 39 KELLESKKTILILHGALGTELESRGCDVSGKLWSAKYLIEDPAAIQTIHEDYIRAGADIVT 98
 + LL+ + L+L GA+ TELE+RGC+++ LWSAK L+E+P I+ +H DY RAGA
 25 Sbjct: 8 RALLDKQDILLDDGAMATELEARGCNLADSLWSAKVLVENPELIREVHLDVYRAGAQC AI 67
 Query: 99 TSTYQATLQGLAQVGVSESTEDLIRLTQVLAQAAREQVWKS LTKEEKSERIYPLISGDV 158
 T++YQAT G A G+ E+Q++ LI +V+LA+ ARE L + ++ + L++G V
 30 Sbjct: 68 TASYQATPAGFAARGLDEAQS KALICKSVELARKAREAY---LAENPQAGTL--LVAGSV 122
 Query: 159 GPYAAFLADGSEYTGLYDIDKQGLKNFHRHRIELLLDEGVDILALETIPNAQEAEALIEL 218
 GPY A+LADGSEY G Y + + FHR R+E LLD G D+LA ET+PN E EAL EL
 Sbjct: 123 GPYGAYLADGSEYRGDYHCSVEAFQAFHRPRVEALLDACADLLACETLPNFSEIEALAE L 182
 35 Query: 219 LAEDFPQVEAYMSFTSQDGK TISDGS AVADLAKAIDVSEPVVALGINCSSPSLVADFLQA 278
 L +P+ A+ SFT +D + +SDG+ + D+ + PQVVALGINC + LQ
 Sbjct: 183 LTA-YPRARAWFSFTLRDSEHLS DGTPLRDVVALLAGYPQVVALGINCIALENTTAALQH 241
 Query: 279 IAEQTNKPLVTYPNSGEVYD GASQSWQSSPDHSHTLLENTSDWQKLG AQVVGCCCRTPA 338
 + T PLV YPNSGE YD S++W +H L + WQ GA+++GGCCRT PA
 40 Sbjct: 242 LHGLTVLPLVVYPNSGEHYDAVSKTWHHGHGHC AQLADYLPQWQAAGARLIGGCCRTTPA 301
 Query: 339 DIADLSA 345
 DIA L A
 45 Sbjct: 302 DIAALKA 308

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

50 A related GBS gene <SEQ ID 8933> and protein <SEQ ID 8934> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 5
 McG: Discrim Score: 5.48
 GvH: Signal Score (-7.5): -2.64
 55 Possible site: 20

-2156-

```
>>> Seems to have an uncleavable N-term signal seq
ALOM program      count: 1 value: -4.78 threshold: 0.0
  INTEGRAL        likelihood = -4.78      Transmembrane      8 - 24 ( 7 - 25)
  PERIPHERAL      likelihood = 2.49       259
modified ALOM score: 1.46
```

*** Reasoning Step: 3

----- Final Results -----

```

bacterial membrane --- Certainty=0.2911(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

ORF01312(412 - 1338 of 1644)
OMNI|NT01EC0303(55 - 357 of 358) conserved hypothetical protein
%Match = 23.8
%Identity = 46.6 %Similarity = 64.3
Matches = 142 Mismatches = 107 Conservative Sub.s = 54

```

288      318      348      378      408      438      468      498
LISQSFC$*FRL*GLLGIAHNVLGFTSVFHLLFSIFAIFTNYVYTRNGDLMGAFKELLESKKTLLHLGALGTELESRGCDVS
::||::|:||:||||:||||::
AWWPVLGWHSIORRELRRCGAGYRLLRCAMVLISLLNPETQNRSONMSQNPLRALDKODILLDGAMATELEARGCNLA
          20           30           40           50           60           70           80

```

528 558 589 618 648 678 708 738
GKLWSAKYLIEDFAAIQTIHEDIYRAGADIVTTSTYQATLQGLAQVGVSESQOTDLIRLTVOQLAKAAREQVWKSLTKEEK
| | | | : : | : : | | | | : : | | : | : : | : : | : : |
DSLWSAKVLVENPELIREVHLDYYRAGAQCATTASYQATPAGFAARGLEDAQS KALIGKSVELARKARE----AYLAEN

 100 110 120 130 140 150

[illegible]

```

1008      1038      1068      1098      1128      1158      1188      1218
EAYMSFTSODGKTI SDGS AVADLAKAIDVSPQVVALGINCSSPSLVADFLQALAEQTNKPLVTYPNSGEVVDGASQSQS
| : ||| : : : ||| : : : : : ||||| : : : : : ||| ||||| : : ||| ||||| : : |||
RAWFSFTLRDSEHLS DGTPLR D VALLAGYPOVVALGINCIALENTAALQHLGLTLVLPVVYPNSCEHYDAVSKTWHH
      250      260      270      280      290      300      310

```

```

1248      1278      1308      1338      1368      1398      1428      1458
SPDHSHTLENTSDWQKLGAVVGGCCRRPADIADLSAHLK*VKYLEEG*GKFDLQSTRKPAWILPNGFCFYLSMT
: | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
HGEHCAQLADYLPQWQAAGARLIGGCCRTTPADIAALKARS
      330      340      350

```

SEQ ID 8934 (GBS381) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 6; MW 42kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 4; MW 66.9kDa).

Example 1912

A DNA sequence (GBSx2021) was identified in *S.agalactiae* <SEQ ID 5927> which encodes the amino acid sequence <SEQ ID 5928>. Analysis of this protein sequence reveals the following:

Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```
bacterial cytoplasm --- Certainty=0.2996(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

-2157-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1913

A DNA sequence (GBSx2022) was identified in *S.agalactiae* <SEQ ID 5929> which encodes the amino acid sequence <SEQ ID 5930>. Analysis of this protein sequence reveals the following:

10 Possible site: 59
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -11.62	Transmembrane	176 - 192 (168 - 198)
INTEGRAL	Likelihood = -11.57	Transmembrane	89 - 105 (80 - 111)
INTEGRAL	Likelihood = -10.03	Transmembrane	337 - 353 (332 - 359)
15 INTEGRAL	Likelihood = -9.87	Transmembrane	292 - 308 (285 - 316)
INTEGRAL	Likelihood = -4.51	Transmembrane	58 - 74 (52 - 78)
INTEGRAL	Likelihood = -3.88	Transmembrane	267 - 283 (267 - 286)
INTEGRAL	Likelihood = -3.08	Transmembrane	125 - 141 (125 - 142)
20 INTEGRAL	Likelihood = -2.13	Transmembrane	212 - 228 (212 - 228)

----- Final Results -----

bacterial membrane ---	Certainty=0.5649(Affirmative)	< succ>
bacterial outside ---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm ---	Certainty=0.0000(Not Clear)	< succ>

25

A related GBS nucleic acid sequence <SEQ ID 9377> which encodes amino acid sequence <SEQ ID 9378> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:CAB12034 GB:Z99105 similar to histidine permease [Bacillus subtilis]
 Identities = 221/384 (57%), Positives = 291/384 (75%), Gaps = 2/384 (0%)

Query: 2 PVTGSFHTYATKFI SPGTGFTVAVLYWICWTVALGTEFLGAAMLQRFNVPVAPAFASF 61
 PVTG+FH TYA K+I PGTGFTVAVLYW+ WTVALG+EF A +LMQRWFP+ W +++
 35 Sbjct: 76 PVTGAFHTYAAKYIGPGTGFTVAVLYWLTWITVALGSEFTAAGLLMQRFNPTSVMWMSAV 135

Query: 62 FALVIRGLNALSVRFPAEAESEFFSSIKVIAIIIFIIILGLGAMFGLVSFEGQHKAILFTHL 121
 FAL IF LNA SV+FFAE+E +FSSIKV+AI++FI+LG AMFG++ +G A + ++
 40 Sbjct: 136 FALFIFLLNAFVSKFFAESEFFWSSIKVLAIVLFILGGSAMFGIIPKGGEAAPMLSNF 195

Query: 122 TANGA-FPNGIVAVVSVMLAVNYAFSGTELGIAAGETDNPKEAVPRAIKTTIGRLVVFF 180
 TA G FPNG V ++ ML+VN+AFSGTELGIAAGE+ +P + +P+AIKTT+ RL +FF
 45 Sbjct: 196 TAEGGLFPNGFVPILMTMLSVNFAPSGTELGIAAGESVDPDKTIPKAIKTTVWRLSLFF 255

Query: 181 VLTIVVLASLLPMKEAGVSTAPFVDVDFKMGIPFTADIMNFVILTALSAAGNSGLYASSR 240
 V TI VL+ L+P+++AGV +PFV VFD++G+P+ ADIMNFVILTALSA NSGLYASSR
 50 Sbjct: 256 VGTIFVLSGLIPIQDAGVIKSPFVAVFDRVGVPAADIMNFVILTALSAANSGLYASSR 315

Query: 241 MLWSLANEGMLSKSVVKINKHGVPMRALLLSMAGAVLSLFSSTIYAADTVYLALVSIAGFA 300
 MLWSL+ E L + K+ G P AL+ SM G +LSL SS++A DTVY+ LVSI+GFA
 55 Sbjct: 316 MLWSLSKEKTLHPTFAKLTSGKTPFNALVFSMIGGILSLLSSVFAPDVTYVVLVSIAGFA 375

Query: 301 VVVVWLAIQVQINERKEFLKE-NQLEDLSYKTPFTPVLPYITIIILLISIVGIAWDSSQ 359
 VVVVW+ I +Q FRK +++ N++ DL Y+TP P +P +L L S+VGIA+D +Q
 Sbjct: 376 VVVVWMGIAASQFMFRKRYIEAGNKVTDLKYRTPLYPFVPIAFLCLASVVGIAFDPNQ 435

55 Query: 360 RAGLYFGVFPFIIFCYIYHKLYKK 383
 R LY GVPF+ CY + ++ +K
 Sbjct: 436 RIALYCCVFPFMAICYAIYVKNRK 459

There is also homology to SEQ ID 4070.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 1914

A DNA sequence (GBSx2023) was identified in *S.agalactiae* <SEQ ID 5931> which encodes the amino acid sequence <SEQ ID 5932>. Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2378(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

There is also homology to SEQ ID 5642.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1915

A DNA sequence (GBSx2024) was identified in *S.agalactiae* <SEQ ID 5933> which encodes the amino acid sequence <SEQ ID 5934>. Analysis of this protein sequence reveals the following:

Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4935(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1916

A DNA sequence (GBSx2025) was identified in *S.agalactiae* <SEQ ID 5935> which encodes the amino acid sequence <SEQ ID 5936>. Analysis of this protein sequence reveals the following:

Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0530(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2159-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1917

A DNA sequence (GBSx2026) was identified in *S.agalactiae* <SEQ ID 5937> which encodes the amino acid sequence <SEQ ID 5938>. Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.0175(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAF63739 GB:AF236863 hypothetical GTP-binding protein
 [Lactococcus lactis]
 Identities = 142/193 (73%), Positives = 165/193 (84%)

20 Query: 6 LNTHNASILLSAANKSHYPQDDLPEVALAGRSNVGKSSFINTLLGRKNLARTSSKPGKTQ 65
 +NT+N +I +SAA+K YP++D PE+ALAGRSNVGKSSFINTLL RKN ARTS +PGKTQ
 Sbjct: 3 ININNLTTTISAASKKQYPENDWPEITALAGRSNVGKSSFINTLLNRKNFARTSGQPGKTQ 62

 Query: 66 LLNFYNIDDKLRFDVPGYGYAKVSKTERAKWGKMIEEYLVTDRNLRVVSLVDFRHDPS 125
 LLNFYNIDD+L FVDVPGYGYA+VSK ER KWKMIIEEYL TR+NL+ VVSLVD RH+PS
25 Sbjct: 63 LLNFYNIDDQLHFVDVPGYGYARVSKKEREKWGKMIEEYLTTRENLKAVVSLVDIRHEPS 122

 Query: 126 ADDIQMYEFLKYYEIPVIVATKADKIPRGKWNKHESIIKKLNFDKDHFI FVSSVDRT 185
 DD+ MYEFLKYY IPVI+VATKADK+PRGKWNKHES IKK + FD D FI+FSS D+T
30 Sbjct: 123 EDDLMMYEFLKYYHIPVILVATKADKVPRGKWNKHESIIKKAMKFDSTDDFIIFSSTDKT 182

 Query: 186 GLDESWDTILSEL 198
 G++E+W IL L
 Sbjct: 183 GIEEAWTAILKYL 195

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5939> which encodes the amino acid sequence <SEQ ID 5940>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0123(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 An alignment of the GAS and GBS proteins is shown below.

 Identities = 167/196 (85%), Positives = 183/196 (93%)

50 Query: 3 EEFLNTHNASILLSAANKSHYPQDDLPEVALAGRSNVGKSSFINTLLGRKNLARTSSKPG 62
 E+ LNTHNASILLSAANKSHYPQDDLPE+ALAGRSNVGKSSFINT+LGRKNLARTSSKPG
 Sbjct: 4 EQVLNTHNASILLSAANKSHYPQDDLPEITALAGRSNVGKSSFINTILGRKNLARTSSKPG 63

 Query: 63 KTQLLNFNIDDKLRFVDVPGYGYAKVSKTERAKWGKMIEEYLVTDRNLRVVSLVDFRH 122
 KTQLLNFN+NIDDKLRFVDVPGYGYAKVSK+ERAKWGKMIEEYL +RDNL R VVSLVD RH
 Sbjct: 64 KTQLLNFFNIDDKLRFVDVPGYGYAKVSKSERAKWGKMIEEYLTSRDNLRVAVVSLVDLRH 123

55 Query: 123 DPSADDIQMYEFLKYYEIPVIVATKADKIPRGKWNKHESIIKKLNFDKDHFI FVSSV 182
 PS +DIQMY+FLKYY+IPVI+VATKADKIPRGKWNKHES +KK LNFDK D FIVFSSV
 Sbjct: 124 APSKEDIQMYDFLKYYDIPVIVVATKADKIPRGKWNKHESVVKALNFDKSDTFIVFSSV 183

-2160-

Query: 183 DRTGLDESWDITLSEL 198
 +R G+D+SWD IL ++
 Sbjct: 184 ERIGDDSWDAILEQV 199

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1918

A DNA sequence (GBSx2027) was identified in *S.galactiae* <SEQ ID 5941> which encodes the amino acid sequence <SEQ ID 5942>. This protein is predicted to be protease ClpX (clpX). Analysis of this
 10 protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2389(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9947> which encodes amino acid sequence <SEQ ID 9948>
 20 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF63738 GB:AF236863 protease ClpX [Lactococcus lactis]
 Identities = 305/395 (77%), Positives = 357/395 (90%), Gaps = 1/395 (0%)

25 Query: 18 NVYCSFCGKSQDEVKKI IAGNGVFCNECVALSQEIIKEELABEVLADLAIEVPKPKELLE 77
 N+ CSFCGKSQD+VKK+TAG+ V+ICNEC+ LS I++EEL EE +++ EV PKE+ +
 Sbjct: 8 NIQCSFCGKSQDDVKKMIAGSDVYICNECIELSTRILEEELKEEQDSEMLEVKTPEMFED 67

30 Query: 78 ILNQYVVGQDRAKRALAVAVNYHKRVSYTESS-DDDVDLQKSNILMIGPTGSGKTFLAQ 136
 LN+YV+GQ++AKRALAVAVNYHKR+++T S +D++LQKSNIL+IGPTGSGKTFLAQ
 Sbjct: 68 HLNEYVIGQEKAKRALAVAVNYHKRINF TASKTAEDIELQKSNILLIGPTGSGKTFLAQ 127

35 Query: 137 TLAKSLNVPFAIADATSLTEAGYVGEDVENILLKLIQAADYNVERAERGIIYVDEIDKIA 196
 TLAKSLNVPFAIADATSLTEAGYVGEDVENILLKL+QA+D+N+ERAERGIIY+DEIDKIA
 Sbjct: 128 TLAKSLNVPFAIADATSLTEAGYVGEDVENILLKLLQASDFNTERAERGIIYIDEIDKIA 187

40 Query: 197 KKGENVISITRDVSGEGVQQALLKII EGTVASVPPQGGGRKHPNQEMIQTINKNIFIVGGA 256
 KK ENVSITRDVSGEGVQQALLKII EGTVASVPPQGGGRKHPNQEMIQT+TKNIFIVGGA
 Sbjct: 188 KKGENVISITRDVSGEGVQQALLKII EGTVASVPPQGGGRKHPNQEMIQIDTKNIFIVGGA 247

45 Query: 257 FDGIEDLVKQRLGEKVIGFGQTSRKIDDNASYMQEIISEDIQKFGLIPEFIGRLPVVAAL 316
 FDGIE++VKQRLGEK+IGFG ++K+ D SYMQEII+EDIQKFGLIPEFIGRLP+VAAL
 Sbjct: 248 FDGIEEIVKQRLGEKIIIGFGANNKLSDEDSYMQEIIAEDIQKFGLIPEFIGRLPIVAAL 307

50 Query: 317 ELLTAEDLVRIITERNALVKQYQTLISYDGVLEFDDQALLAIADKAIERKTGARGLRS 376
 E LT EDL++ILTEP+NAL+KQY+ LL +D VELEF AL+AIA KAIERKTGARGLRS
 Sbjct: 308 ERLTEEDLIQILTEPKNALIKQYQQLLLFDNVELEFQD GALMAIAKKAIERKTGARGLRS 367

Query: 377 IIEETMLDIMFEIPSQEDVTKVRIITKA AVEGTDKP 411
 IIEE M+DIMFE+PS E++TKV IT+A V+G +P
 Sbjct: 368 IIEEVMDIMFEVPSHEETKVIITEAVVDGKAEP 402

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5943> which encodes the amino acid sequence <SEQ ID 5944>. Analysis of this protein sequence reveals the following:

55 Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-2161-

bacterial cytoplasm --- Certainty=0.2711(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 378/409 (92%), Positives = 393/409 (95%), Gaps = 1/409 (0%)

```

Query: 9  MAGNRNNDMMNVYCSFCGKSQDEVKKIIAGNGVFICNECVALSQEIIKEELABEVLADLAE 68
          MAG+R ND+ VYCSFCGKSQD+VKKIIAGN VFICNECVALSQEIIKEELABEVLADL E
10  Sbjet: 1  MAGSRTNDIKVYCSFCGKSQDDVKIIAGNVFICNECVALSQEIIKEELABEVLADLTE 60

Query: 69  VPKPKELLEILNQYVVGQDRAKRALAVAVNHYKRVSYTES-SDDVDLQKSNILMIGPT 127
          VPKPKELL++LNQYVVGQDRAKRAL+VAVNHYKRVSTES DDDVDLQKSNILMIGPT
15  Sbjet: 61  VPKPKELLDVLNQYVVGQDRAKRALSAVVNHKRVSTESRDDDVLQKSNILMIGPT 120

Query: 128  GSGKTFLAQTAKSLNVPFAIADATSLTEAGYVGEDVENILLKLIQAADYNVERAERGII 187
          GSGKTFLAQTAKSLNVPFAIADATSLTEAGYVGEDVENILLKLIQAADYNVERAERGII
20  Sbjet: 121  GSGKTFLAQTAKSLNVPFAIADATSLTEAGYVGEDVENILLKLIQAADYNVERAERGII 180

Query: 188  YVDEIDKIAKKGENVSITRDVSGEGVQALLKIIEGTVASVPPQGRKHPNQEMIQTNTK 247
          YVDEIDKIAKKGENVSITRDVSGEGVQALLKIIEGTVASVPPQGRKHPNQEMIQT+TK
25  Sbjet: 181  YVDEIDKIAKKGENVSITRDVSGEGVQALLKIIEGTVASVPPQGRKHPNQEMIQTIDTK 240

Query: 248  NILFIVGGAFDGIEDLVKQRLGEKVIGFGQTSRKIDDNASYMQEIISEDIQKFGLIPEFI 307
          NILFIVGGAFDGIIE++VKQRLGEKVIGFGQ SRKIDDNASYMQEIISEDIQKFGLIPEFI
30  Sbjet: 241  NILFIVGGAFDGIIEIVKQRLGEKVIGFGQNSRKIDDNASYMQEIISEDIQKFGLIPEFI 300

Query: 308  GRLPVVALELLTAEDLVRLITEPRNALVKQYQTLLSYDGVLEFDQDALLAIADKAIER 367
          GRLPVVALE L DL++ILTEPRNALVKQYQ LLSYDGVLEL FD++AL AIA+KAIER
35  Sbjet: 301  GRLPVVALEQLNTSDLIQILTEPRNALVKQYQALLSYDGVLELAFDKEALEATANKAIER 360

Query: 368  KTGARGLRSIIEETMLDIMFEIPSQEDVTKVRIITKAAVEGTDKPVLETA 416
          KTGARGLRSIIEETMLDIMFEIPSQEDVTKVRIITKAAVEG KPVLETA
40  Sbjet: 361  KTGARGLRSIIEETMLDIMFEIPSQEDVTKVRIITKAAVEGSKPVLETA 409
  
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1919

A DNA sequence (GBSx2028) was identified in *S.agalactiae* <SEQ ID 5945> which encodes the amino acid sequence <SEQ ID 5946>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

45 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

50 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1920

A DNA sequence (GBSx2029) was identified in *S.agalactiae* <SEQ ID 5947> which encodes the amino acid sequence <SEQ ID 5948>. Analysis of this protein sequence reveals the following:

55

-2162-

Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.4029(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9949> which encodes amino acid sequence <SEQ ID 9950>
 10 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC33872 GB:AF055727 dihydrofolate reductase [Streptococcus pneumoniae]
 Identities = 83/162 (51%), Positives = 118/162 (72%), Gaps = 1/162 (0%)

15 Query: 25 MTKQIIAIWAEDHLIGVNGGLPWRLPKELHHFKETTMGQALLMGRKTFDGMNRRVLP 84
 MTK+I+AIWA+DE+ LIG LPW LP EL HFKETT+ A+LMGR TFDGM RR+LP
 Sbjct: 1 MTKKIVAIWAQDEGLIGKENRLPWHLPAELQHFKETTTLNHAILMGRVTFDGMGRLLLPK 60

20 Query: 85 RETIILTKDEQFQADGVTVLNSVEQVIKWFQEHNKTLFIVGGASIYKAFLPYCEAIKTK 144
 RET+ILT++ + + DGV V+ V+ W+Q+ K L+I+GG I++AF PY + +I T
 Sbjct: 61 RETILITRNPEEKIDGVATFQDVQSVLDWYQDQEKNLIIIGGKQIFQAFEPYLDIVTH 120

 Query: 145 VHGFKFGDTYFP-DVNLSEFKVISRDYFEKDEQNAHAFTVTY 185
 +H + +GDTYFP +++LS F+ +S ++ KDE+N + FT+ Y

25 Sbjct: 121 IHARVEGDTYFPEELDLSLFETVSSKPYAKDEKNPYDFTIQY 162

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5949> which encodes the amino acid
 sequence <SEQ ID 5950>. Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.1214(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 82/160 (51%), Positives = 119/160 (74%)

40 Query: 25 MTKQIIAIWAEDHLIGVNGGLPWRLPKELHHFKETTMGQALLMGRKTFDGMNRRVLP 84
 MTK+IIAIWAED LIG+ G LPW LPKEL HFK+IT+ QA+LMGR TF+GMN + LP
 Sbjct: 1 MTKEIIAIWAED EAGLIGTAGKLPWYLPKELEHFKTTTLHQAAILMGRVTFEGMNCKRLPQ 60

45 Query: 85 RETIILTKDEQFQADGVTVLNSVEQVIKWFQEHNKTLFIVGGASIYKAFLPYCEAIKTK 144
 R+T+++T++ +Q D V + S+E+V++W+ +KTL+I+GG + +AF Y + IIKT
 Sbjct: 61 RQTLVMTRNRDYQVDEVLTMTSIEKVLEWYHAQDKTLYIIGGNKVLEAFNGYFDRIIKTV 120

 Query: 145 VHGFKFGDTYFPDVNLSEFKVISRDYFEKDEQNAHAFTVT 184
 +H +FKGDTY P+++ S F S+ ++ +D +N + FTVT

50 Sbjct: 121 IHHRFKGDTYRENLD FSHFTQESQTFYARDAKNPYDFTVT 160

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 1921

55 A DNA sequence (GBSx2030) was identified in *S.agalactiae* <SEQ ID 5951> which encodes the amino
 acid sequence <SEQ ID 5952>. Analysis of this protein sequence reveals the following:

Possible site: 45

-2163-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1577(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAA25221 GB:M33770 thymidylate synthase (EC 2.1.1.45)
 [Lactococcus lactis]
 Identities = 215/280 (76%), Positives = 245/280 (86%), Gaps = 2/280 (0%)

Query: 1 MTKADLLFKDNITKIMSEGVFSEQARPRYKNGEMANSKYITGAFAEYDLSKGEFPITTLR 60
 MT AD +FK NI I+ GVFSE ARP+YK+G+MANSKY+TG+F YDL KGEFPITTLR

15 Sbjct: 1 MTYADQVFKQNIQNILDNGVFSENARPKYKDGQMANSKYVTGSFVTYDQLKGEFPITTLR 60

Query: 61 PIPIKSAIKEIFWIYQDQTNDLAVLNDKYGVTYWNDWEVGHTGTIGQRYGAVVKKHNIIS 120
 PIPIKSAIKE+ WIYQDQT++L+VL +KYGV YW +W +G GTIGQRYGA VKK+NII

20 Sbjct: 61 PIPIKSAIKELMWIYQDQTSLSVLEEKYGVKYWGEWGIGD-GTIGQRYGATVKKYNIIG 119

Query: 121 KLLKQLEDNPNWRRNVISLWDYEAFFEETEGLLPCAFQTMFDVRRV-NGELYLDATLTQRS 179
 KLL+ L NPWNRN+I+LW YE FEETEGLLPCAFQTMFDVRR +G++YLDATL QRS

Sbjct: 120 KLLEGLAKNPWNRNIIINLWQYEDFEETEGLLPCAFQTMFDVRRKDGQIYLDATLIQRS 179

25 Query: 180 NDMLVAHHINAMQYVALQMMIAKHFGWRVGKFFYFINNLHIYDNQFEQAQELLKRQPSEC 239
 NDMLVAHHINAMQYVALQMMIAKHF W+VGKFFYF+NNLHIYDNQFEQA EL+KR SE

Sbjct: 180 NDMLVAHHINAMQYVALQMMIAKHFSWKVGKFFYFVNNLHIYDNQFEQANELMKRTASEK 239

30 Query: 240 NPKLVNLNVPDGTDFFDIKPDDFALVDYDPIKPQLRFDLAI 279
 P+LVLNVPDGT+FFDIKP+DF LVDY+P+KPQL+FDLAI

Sbjct: 240 EPRLVLNVPDGTNFFDIKPEDFELVDYBPVKPQLKFDLAI 279

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5953> which encodes the amino acid sequence <SEQ ID 5954>. Analysis of this protein sequence reveals the following:

35 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.3131(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

45 Identities = 227/279 (81%), Positives = 251/279 (89%)

Query: 1 MTKADLLFKDNITKIMSEGVFSEQARPRYKNGEMANSKYITGAFAEYDLSKGEFPITTLR 60
 MTKAD +FK NI KI++EG SEQARP+YK+G A+SKYITGAFAEYDL+KGEFPITTLR

50 Sbjct: 9 MTKADQIFKANIQKIIINEGSLSEQARPKYKDGRTAHSKYITGAFAEYDLAKGEFPITTLR 68

Query: 61 PIPIKSAIKEIFWIYQDQTNDLAVLNDKYGVTYWNDWEVGHTGTIGQRYGAVVKKHNIIS 120
 PIPIKSAIKE+FWIYQDQ+N L VL KY V YWN+WEV T TIGQRYGAVVKKH+IIS

55 Sbjct: 69 PIPIKSAIKELFWIYQDQSNLSLDVLEAKYNVHYWNEWEVDQTRTTIGQRYGAVVKKHDIIS 128

Query: 121 KLLKQLEDNPNWRRNVISLWDYEAFFEETEGLLPCAFQTMFDVRRVNGELYLDATLTQRSN 180
 K+LKQL +NPWNRN+VVISLWDYEAFFEET+GLLPCAFQ MFDVRRV +LYLDA+LTQRSN

60 Sbjct: 129 KILKQLAENPNWRRNVISLWDYEAFEETKGLLPCAFQIMFDVRRVGEDLYLDASLTQRSN 188

Query: 181 DMLVAHHINAMQYVALQMMIAKHFGWRVGKFFYFINNLHIYDNQFEQAQELLKRQPSECN 240
 D+LVAHHINAMQYVALQMMIAKHFGW++GKFFYF+NNLHIYDNQF+QAQELLKRQP

Sbjct: 189 DILVAHHINAMQYVALQMMIAKHFGWKIGKFFYFVNNLHIYDNQFDQAQELLKRQPVASQ 248

Query: 241 PKLVNLNVPDGTDFFDIKPDDFALVDYDPIKPQLRFDLAI 279
 PKLVNLNVPD T+FFDIKPDDF L +YDP+KPQL FDLAI

Sbjct: 249 PKLVNLNVPDRNTNFFDIKPDDFELQNYDPVKPQLHFDLAI 287

-2164-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1922

- 5 A DNA sequence (GBSx2031) was identified in *S.agalactiae* <SEQ ID 5955> which encodes the amino acid sequence <SEQ ID 5956>. This protein is predicted to be HMG-CoA synthase. Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

10

```
----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0816(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

15

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5957> which encodes the amino acid sequence <SEQ ID 5958>. Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have no N-terminal signal sequence

20

```
----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1670(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

25

An alignment of the GAS and GBS proteins is shown below.

Identities = 260/385 (67%), Positives = 325/385 (83%)

30

```
Query: 36 MKIGIDKIGFATSQYVLEMTDLAIARQVDPEKFSKGLLLDSLSITPVTEIDIVTLAASAAN 95
      M IGIDKIGFATSQYVL++ DLA+ARQVDP KFS+GLL++S S+ P+TEDI+TLAASAA+
Sbjct: 14 MTIGIDKIGFATSQYVLKLEDLALARQVDPKFSQGLLIESFSVAPITEDIITLAASAAD 73
```

35

```
Query: 96 DILSDEDEKETIDMVIVATESSIDQSKAASVYVHQLLEIQPFARSFEMKEACYSATAALDY 155
      IL+DED+ IDMTI+ATESS DQSKA+++YVH L+ IQPFARSFE+K+ACYSATAALDY
Sbjct: 74 QILTDEDRAKIDMVILATESSTDQSKASAIYVHHLVGIQPFARSFEVQKQACYSATAALDY 133
```

40

```
Query: 156 AKLHVEKHPDSKVLVIASDIKYGKSTGESTQCGAGSIAMLSIQNPSILELKEDHLAQTR 215
      AKLHV PDS+VLVIASDIA+YG+ S GESTQG+GSIA+L++ NP IL L ED++AQTR
Sbjct: 134 AKLHVASKPDSRVLVIASDIARYGVGSPGESTQSGSIALLVNTANPRILALNEDNVAQTR 193
```

45

```
Query: 216 DIMDFWRPNYSVPYVNGMFSTKQYLDMLKTTWKVYQKRFTSLSDYAAFCFHIPFPKLA 275
      DIMDFWRPNYS PYV+G++STKQYL+ L+TTW+ YQKR N LSD AA CFHIPPFPKLA
Sbjct: 194 DIMDFWRPNYSFTPYVDGIYSTKQYLNCLLETTWQAYQKRENQLSLAAVCFHIPPFPKLA 253
```

50

```
Query: 276 LKGFNKLNNLDEQKKAELQENFEHSITYSKKIGNCYTGSLLYLGLLSLENSQNLKAGD 335
      LKG N I+DN + + + +L E F+ SI+YSK+IGN YTGSLYLGLLSLENS+ L++GD
Sbjct: 254 LKGLNNIMDNTVPEHREKLIKAFQASISYSKQIGNIYTGSLYLGLLSLENSKVLQSGD 313
```

55

```
Query: 336 QIAFFSYSGGAVAEIFTGQLVDGYQNKQLQSDRMDQLNKRQKITVTEYEKLFPEKTILDEN 395
      +I FFSYSGGAV+E ++GQLV GY L ++R L++R +++V++YE LF+E+ LD+N
Sbjct: 314 KIGFFSYSGGAVSEFYSGQLVAGYDKMLMTNRQALLDQRTLSVSKYEDLFYEQVQLDDN 373
```

```
Query: 396 GNANFNTYRTGTFSLDICEHQRIY 420
      GNANF+ Y TG F+L +I EH+RIY
```

```
Sbjct: 374 GNANFDIYLTGKFALTAIKEHRRY 398
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2165-

Example 1923

A DNA sequence (GBSx2032) was identified in *S.agalactiae* <SEQ ID 5959> which encodes the amino acid sequence <SEQ ID 5960>. This protein is predicted to be HMG-CoA reductase (mvaA). Analysis of this protein sequence reveals the following:

```

5   Possible site: 50
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.49    Transmembrane  348 - 364 ( 348 - 364)
      INTEGRAL    Likelihood = -1.33    Transmembrane  53 - 69 ( 53 - 69)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15  The protein has homology with the following sequences in the GENPEPT database.
   >GP:AAG02454 GB:AF290098 HMG-CoA reductase [Streptococcus pneumoniae]
      Identities = 266/421 (63%), Positives = 343/421 (81%), Gaps = 3/421 (0%)

20  Query: 3   KISWTGFSKKSPEERIHYLEEQDFLADSSLEIVTNQDLLSLSLANQMAENVIGRIALPFS 62
      KISW GFSKKS +ER+ L+ Q L+ + + +S+++A+Q++ENV+G +LP+S
      Sbjct: 2   KISWNGFSKKSQYERLELLKAQALLSPERQASLEKDEQMSVTVADQLSENVVGTFSLPYS 61

      Query: 63   LVPDVLVNGKVYQVPYVTEEPSVVAASFAAKIIKRSGGFLTTHNRMIGQVALYDVQD 122
      LVP+VLVNG+ Y VFYVTEEPSVVAAS+A+KIIKR+GGF VH R+MIGQVALY V +
25  Sbjct: 62   LVPEVLVNGQGYTVFYVTEEPSVVAASVASKIIKRAGGFTAQVHQRMIGQVALYQVAN 121

      Query: 123  SQHTKESILNQKQQLLEIANAAHPSIVKRGGCACDLTIEI---KEDFLIVYLMVDTEAM 179
      + +E I ++K +LLE+AN A+PSIVKRGGGA DL +E + DFL+VY+ VDT+EAM
30  Sbjct: 122  PKLAQEKIASKKAELLELANQAYPSIVKRGGGARDLHVEQIKGEPDFLVVYIHVDTQEAM 181

      Query: 180  GANMVMNTMEALSSPLEDISKGSLSMSILSNYATESLVTATCRVDLRFLSRQKEEAIKLA 239
      GANM+NTM+EAL LE++S+G+SLM ILSNYAT+SLVTA+CR+ R+LSRQK++ ++A
      Sbjct: 182  GANMLNTMLEALKPVLEELSQQQSLMGILSNYATDSLVTASCRIFRYLSRQKQDQGREIA 241

35  Query: 240  QKMTMASQLAQVDPYRASTHNKGIFNGIDAIVLATGNDWRAIEAGAHTYAVKDGQYRGLS 299
      +K+ +ASQ AQ DPYRA+THNKGIFNGIDAI++ATGNDWRAIEAGAH +A +DG+Y+GLS
      Sbjct: 242  EKIALASQFAQADPYRAATHNKGIFNGIDAILIATGNDWRAIEAGAHAFASRDGRYQGLS 301

      Query: 300  RWSYKVDNDCLEGTLTLEMPVATKGGSIGINPSVHLAHDLLGRPNAKELASIIISIGLAQ 359
      W+ ++ L G +TLEMPVATKGGSIG+NP V L+HDLG P+A+ELA II+SIGLAQ
40  Sbjct: 302  CWILDLEREELVGEMTLMPVATKGGSIGLNPRVALSHDLLGNPSARELAQIIIVSIGLAQ 361

      Query: 360  NFAALKALVSTGIQAGHMKLQAKSLALLAGAKEEQISEVVKQLLDSKHMNLETAQKIVNKL 420
      NFAALKALVSTGIQ GHMKLQAKSLALLAGA E +++ +V++L+ K NLETAQ+ + L
45  Sbjct: 362  NFAALKALVSTGIQQGHMKLQAKSLALLAGASEVAPLVERLISDKTFNLETAQRYLENL 422

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5961> which encodes the amino acid sequence <SEQ ID 5962>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 31
   >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3929(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
55  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 257/422 (60%), Positives = 330/422 (77%)

60  Query: 2   TKISWTGFSKKSPEERIHYLEEQDFLADSSLEIVTNQDLLSLSLANQMAENVIGRIALPF 61
      T ++W+GFSKK+ EER+ +E+ L +L+ + LL+ ANQM ENV+GR+ALPF

```

-2166-

Sbjct: 4 TNLNWSGFSKKTFEERLQLIEKFKLLNAENLNQLKTDVLLPIQTANQMTENVIGRLALPF 63

Query: 62 SLVPDVLVNGKVYPVYVTEEPSVVAASFAAKIKRSGGFLTTHVNRKMIGQVALYDVQ 121
S+ PD LVNG YQ+P+VTEEPSVVAASFAAK+IKRSGGF NR+MIGQ+ LYD+

5 Sbjct: 64 SIAPDFLVNGSTYQMPFVTEEPSVVAASFAAKIKRSGGFKAQTLNRQMIGQIVLYDID 123

Query: 122 DSQHTKESILNQKQQLLEIANAAHPSIVKRGGGACDLTIEIKEDFLIVYLMVDTKEAMGA 181
+ K +IL++ ++L+ +AN A+PSIVKRGGGA + +E K +FLI YL VDT+EAMGA

10 Sbjct: 124 QIDNAKAAAILHKTCKLIALANKAYPSIVKRGGGARTIHLEEKGEFLIFYLTVDTEAMGA 183

Query: 182 NMVNTMMEALSSPLEDISKGKSLMSILSNYATESLVTATCRVDLRFLSRQKEEAIKLAQK 241
NMVNTMMEAL L +SKG LM+ILSNYATESLVT +C + +R L K ++++LAQK

Sbjct: 184 NMVNTMMEALVPDLTRLKSGHCLMAILSNYATESLVTTSCEIPVRLDHDKTKSLQLAQK 243

15 Query: 242 MTMASQLAQVDPYRASTHNGGIFNGIDAIVLATGNDWRAIEAGAHTYAVKDGQYRGLSRW 301
+ +AS+LAQVDPYRA+THNGGIFNGIDA+V+ATGNDWRAIEAGAH YA ++G Y+GLS+W

Sbjct: 244 IELASRLAQVDPYRATTHNGGIFNGIDAVVIATGNDWRAIEAGAHAYASRNGSYQGLSQW 303

20 Query: 302 SYKVDNCLBEGTTLTLPMPVATKGGSIGINPSVHLAHDLLGRPNAKELASITLSIGLAQNF 361
+ D L G +TLPMP+A+KGGSIG+NP+V +AHDLL +P+AK LA +I S+GLAQNF

Sbjct: 304 HFDQDKQVLLGQMTLPMPIASKGSGIGLNPTVSHADLLNQPDATLAQLIASVGLAQNF 363

Query: 362 AALKALVSTGIQAGHMKLOAKSLALLAGAKEEQISEVVKQLLDSKHMNLETAQKIVNKLIT 421
AALKAL S+GIQAGHMKL AKSLALLAGA +++I+ +V LL K +NLE A +++L

25 Sbjct: 364 AALKALTSSGIQAGHMKLHAKSLALLAGATQDETAPLVNALLADKPINLEKAHFYLSQLR 423

Query: 422 KS 423
+S

30 Sbjct: 424 QS 425

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1924

A DNA sequence (GBSx2033) was identified in *S.galactiae* <SEQ ID 5963> which encodes the amino acid sequence <SEQ ID 5964>. Analysis of this protein sequence reveals the following:

Possible site: 52
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2355(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5965> which encodes the amino acid sequence <SEQ ID 5966>. Analysis of this protein sequence reveals the following:

Possible site: 53
>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2687(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 An alignment of the GAS and GBS proteins is shown below.

Identities = 76/138 (55%), Positives = 100/138 (72%), Gaps = 2/138 (1%)

Query: 7 PKWEELPELDLYLDQVLLYVNLINPKTITNDKLLTASMINNVYKHNYISKPIKKKYNR 66
P W++LP+LDLYLDQVLLYVNQ + ++++K LTASMINNVYKH Y++KPIKKKY ++

-2167-

Sbjct: 7 PYWKDLPLDLVLYDQVLLYVNQCTDFSEVSDNKSILTASMINNYVKGHYVTKPIKKYQKQ 66

Query: 67 QVARLIVTTAPKQVFQAIQEISQTLLELLTADNHSEEAAYNGFAACMNKEE--VHDLPPVVIS 124
Q+ARLI I+ FK VF IQ+IS+ LE L A SE YN F C N++ D+PP+V

Sbjct: 67 QLARLIAISLFTVFPIQDISRVLEELQAQADSESLYNTFTVCWNQKAPIEDDIPPIVQV 126

Query: 125 ACQTLNLYQETQKLVLEL 142
ACQT+ Y +T L+ E+

Sbjct: 127 ACQTVKDYHKTIYLLQEV 144

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1925

A DNA sequence (GBSx2034) was identified in *S.galactiae* <SEQ ID 5967> which encodes the amino acid sequence <SEQ ID 5968>. This protein is predicted to be hemolysin iii. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.08	Transmembrane	142 - 158 (140 - 165)
INTEGRAL	Likelihood = -6.79	Transmembrane	26 - 42 (19 - 44)
INTEGRAL	Likelihood = -5.63	Transmembrane	200 - 216 (196 - 217)
INTEGRAL	Likelihood = -5.41	Transmembrane	104 - 120 (102 - 121)
INTEGRAL	Likelihood = -3.98	Transmembrane	51 - 67 (49 - 69)
INTEGRAL	Likelihood = -1.86	Transmembrane	172 - 188 (169 - 188)

----- Final Results -----

bacterial membrane	---	Certainty=0.4630(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 9951> which encodes amino acid sequence <SEQ ID 9952> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA58877 GB:X84058 novel hemolytic factor [Bacillus cereus]

Identities = 79/204 (38%), Positives = 132/204 (63%), Gaps = 4/204 (1%)

Query: 17 EELANSITHAVGALLMLLIPITAVYSHNHFGLOALGTSIFVTSFLFMLFSSSIYHSMT 76
EE+AN+ITH +GA+L + L I +++ H A + +++ S+FL++L S++ HS+

Sbjct: 14 EEIANATHTGIGAILSIPALIILIIHASKHGTASAVVAPTIVYGVSMFLLYLFSTLLHSIH 73

Query: 77 YNSLQKYVLRMIDHSMIYIAIAGSYTPVALSLIGGWLYLIIFLQWGITLFGILYKIFAP 136
+ ++K + ++DHS IY+ IAG+YTP L + G LG+ ++ + W + + GI++KIF

Sbjct: 74 HPKVEK-LFTILDHSAIYLLIAGTYTPFLLITLRGPLGWILLAIWFLAIGGIIFKIFV 132

Query: 137 KINDKFSVLVLYLIMGWLVIIF-IPPAITKTGPAPWGLLLAGGICYTIGALFYA-RKRPYD 194
+ K S + Y+IMGWL+I I P TG F LLLAGGI Y++GA+F+ K P++

Sbjct: 133 RRFKASTLCYIIMGWLIIVAIKPLYENLTGHGF-SLLLAGGILYSVGAIFFLWEKLPFN 191

Query: 195 HMIWHLFILLASILQYIGIVYFML 218
H IWHLF+L S + + +++++L

Sbjct: 192 HAIWHLFVLGGSAMFFCVLFYVL 215

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5969> which encodes the amino acid sequence <SEQ ID 5970>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.51	Transmembrane	144 - 160 (138 - 163)
INTEGRAL	Likelihood = -9.87	Transmembrane	49 - 65 (45 - 71)

-2168-

INTEGRAL Likelihood = -7.11 Transmembrane 198 - 214 (193 - 215)
 INTEGRAL Likelihood = -6.16 Transmembrane 102 - 118 (100 - 120)
 INTEGRAL Likelihood = -2.97 Transmembrane 20 - 36 (20 - 41)
 INTEGRAL Likelihood = -1.01 Transmembrane 167 - 183 (167 - 185)

----- Final Results -----

bacterial membrane --- Certainty=0.5203 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA58877 GB:X84058 novel hemolytic factor [Bacillus cereus]

Identities = 82/204 (40%), Positives = 128/204 (62%), Gaps = 4/204 (1%)

Query: 15 EEVANSVTHAIGAFAMLLPISASYAYQTYDLKAAIGISIFVISLFLMFLSSTIYHSMA 74
 EE+AN++TH IGA + L I +A + A + +++ +S+FL++L ST+ HS+
 Sbjct: 14 EEIANAITHGIGAILSIPALIILIIHASKHGTASAVVAFTVYGVSMFLLYLFTLLHSIH 73

Query: 75 YGSVHKYILRIIDHSMIYIAIAGSYTPVALSLVSGWLGYIIIVLQWGITLFGILYKIFAK 134
 + V K + I+DHS IY+ IAG+YTP L + G LG+ ++ + W + + GI++KIF
 Sbjct: 74 HPKVEK-LFTILDHSAIYLLIAGTYTPFLILITLRLGFLGWTLLAIWTLAIGGIIFKIFFV 132

Query: 135 RINEKFSMLMYIVMGWL-VVFILPVIIQKTSIAFGLMLFGGLSYTIGAVFYA-KKRPYF 192
 R K S + YI+MGWL +V I P+ T F LL L GG+ Y++GA+F+ +K P+
 Sbjct: 133 RRFIKASTLCYIIMGWLIIVAIKPLYENLTGHGFSLL-LAGGILYSVGAIFFLWEKLPFN 191

Query: 193 HMIWHLFILLASALQFIATFFML 216
 H IWHLF+L SA+ F + F++L

Sbjct: 192 HAIWHLFVLGGSAMMFFCVLFYVL 215

An alignment of the GAS and GBS proteins is shown below.

Identities = 153/213 (71%), Positives = 181/213 (84%)

Query: 6 SIKLSPQLSFGHEELANSITHAVGALLMLLIPITAVYSHNHFGLOALGTSIFVTSLFLM 65
 + K S LSF EE+ANS+THA+GA MLILLPT+A Y++ + L+AA+G SIFV SLFLM
 Sbjct: 4 TPKQSLPLSFSEEVANSVTHAIGAFAMLLPISASYAYQTYDLKAAIGISIFVISLFLM 63

Query: 66 FLSSSIYHSMYNSLQXYVLRMIDHSMIYIAIAGSYTPVALSLIGGWLGYYLIIFLQWGIT 125
 FLSS+IYHSM Y S+ KY+LR+IDHSMIYIAIAGSYTPVALSL+ GWLGYY+II LQWGIT
 Sbjct: 64 FLSSSIYHSMAYGSVHKYILRIIDHSMIYIAIAGSYTPVALSLVSGWLGYIIIVLQWGIT 123

Query: 126 LFGILYKIFAPKINDKFSLVLYLIMGWLVIFIFPAIITKTGPAFWGILLAGGICYTIGAL 185
 LFGILYKIFA +IN+KFSL+LY++MGWLV+FI P II KT AF L+L GG+ YTIGA+
 Sbjct: 124 LFGILYKIFAKRINEKFSMLMYIVMGWL-VVFILPVIIQKTSIAFGLMLFGGLSYTIGAV 183

Query: 186 FYARKRPYDHMIWHLFILLASILQYIGIVYFML 218
 FYA+KRPY HMIWHLFILLAS LQ+I I +FML
 Sbjct: 184 FYAKRPYFHMIIWHLFILLASALQFIATFFML 216

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1926

A DNA sequence (GBSx2035) was identified in *S.agalactiae* <SEQ ID 5971> which encodes the amino acid sequence <SEQ ID 5972>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3641 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-2169-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12492 GB:Z99107 similar to hypothetical proteins [Bacillus subtilis]
Identities = 81/302 (26%), Positives = 157/302 (51%), Gaps = 10/302 (3%)

Query: 1 MKSAYIFFNPKSGKDEQALAKEVKSYLEHDFQDDY-VRIITPSSVEEAVAKKASEDH 59
MK A I +NP SG++ + K+ + +++ Q Y + +A AK+A+
Sbjct: 1 MKRARIINYNTSGRE---IFKKHLAQVLQKFEQAGYETSTHATTACAGDATHAAKEAALRE 57

Query: 60 IDLVIPLGGDGTINKICGGVYAGGAYPTIGLVPAGTVNNFSKALNIPQERNL-ALENLLN 118
DL+I GGDGTIN++ G+ PT+G++P GT N+F++AL IP+E L A + ++N
Sbjct: 58 FDLIIAAGDGTINEVNGLAPLDNRPTLGVIPVGTITNDFARALGIPREDILKAADTVIN 117

Query: 119 GHVKSVDICKVNDYMISSLTGLLADIAANVTSEMKRKLGPFAGLGDAYRILKRNRSYS 178
G + +DI +VN Y I+ G L ++ +V S++K LG A+ +L R
Sbjct: 118 GVARPIDIGQVNGQYFNIAGGRLTELTYDVPSKLTMLGQLAYYLKGMEMPLSLRPTE 177

Query: 179 TILAYDNNVRSRLRLLITMTNSIAGMPAFSPEATIDDGLFRVYTMETIHFFKLLHLR 238
+ + YD + L L+T+INS+ G +P+++++DG+F + ++ + + +
Sbjct: 178 VEIEVDGKLFQGEIMLFLVTLTNSVGGFEKLAPDSSLNDGMFDMILKKNLAEFIRVAT 237

Query: 239 QFRKGDFSQAKEIKHFHTNNLTISTFKRKKSAIPKVRIDGDPGDQLPVKVEVIPKALKFI 298
+G+ + I + N + ++ ++ ++ +DG+ G LP + + + + +
Sbjct: 238 MALRGEHINDQHIIYTKANRVKVNSEKM-----QLNLDGEYGGMLPGEFVNLYRHHV 292

Query: 299 IP 300
+P
Sbjct: 293 MP 294

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5119> which encodes the amino acid sequence <SEQ ID 5120>. Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4258(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 172/300 (57%), Positives = 229/300 (76%)

Query: 1 MKSAYIFFNPKSGKDEQALAKEVKSYLEHDFQDDYVRIITPSSVEEAVAKKASEDH 60
MK+ IF+NP SGK E LA++VK Y +H F +D V++ITP ++A LAK+A++D I
Sbjct: 1 MKTVRIFYNPNSGKKESQLARQVKDYFCQHGFSEDSVKVITPKDADQAFQLAKQAADKI 60

Query: 61 DLVIPLGGDGTINKICGGVYAGGAYPTIGLVPAGTVNNFSKALNIPQERNLALENLLNGH 120
DLVIPLGGDGT+NKI GG+Y GCA+ IGLVP+GTVNNF+KA++IP + AL+ +L G
Sbjct: 61 DLVIPLGGDGTLNKIIGGIYEGGAHCLIGLVPSGTVNNFAKAMHIPLQITEALDTILTQ 120

Query: 121 VKSVDICKVNDYMISSLTGLLADIAANVTSEMKRKLGPFAGLGDAYRILKRNRSYSIT 180
+K VDICK N YMISSLTGLLADIAA+VT+E KR+ GP AFL D+ RILKRNRSY+I+
Sbjct: 121 IKQVDICKANQQYMISSLTGLLADIAADVTAEEKRRFGPLAFLKDSIRILKRNRSYALS 180

Query: 181 LAYDNNVRSRLRLLITMTNSIAGMPAFSPEATIDDGLFRVYTMETIHFFKLLHLRQF 240
L N+ L+T+ LLITMTN+IAG P+FSP A DDG F+VYTM+ + FFK L H+ F
Sbjct: 181 LISHNRHILKTKFLLITMTNTIAGFSPFSGAQADDGYFQVYTMKKVSFFKFLWHINDF 240

Query: 241 RKGDFSQAKEIKHFHTNNLTISTFKRKKSAIPKVRIDGDPGDQLPVKVEVIPKALKFIIP 300
++GDFS+A+EI HF N L++ +K++ +P+ RIDGD D LP+++++IPKA+ I+P
Sbjct: 241 KQGDFSQAEIISHFQANTLSLLPQAKKQAILPRTRIDGDKSDYLPQLDIIPKAVSIIVP 300

-2170-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1927

A DNA sequence (GBSx2036) was identified in *S.agalactiae* <SEQ ID 5973> which encodes the amino acid sequence <SEQ ID 5974>. Analysis of this protein sequence reveals the following:

Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.3628(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:BAB10885 GB:AB010693 gene_id:K21C13.21~pir||T04769~strong
 similarity to unknown protein [Arabidopsis thaliana]
 Identities = 85/291 (29%), Positives = 150/291 (51%), Gaps = 28/291 (9%)

20 Query: 10 DQEWEPVVESGRYHMIVGEFCPYAQRPOIARQLLGLDKHISISFVDDV----- 57
 D + + P ESGRYH+ + CP+A R ++ GLD+ I+ S V +
Sbjct: 29 DPDSQPPAESGRVHLYISYACPWACRCLSYLKI KGLDEAITFSSVHAIWGRKTKETDDHRG 88

Query: 58 ----PSDIGLIFSQFEQVTGAKSLRDIYHLTDPTYQGPYTIPIILIDKTDNRIVCKESADI 113
 SD L ++P+ + GAKS+R++Y + P Y+G YT+P+L DK +V ES+++
25 Sbjct: 89 WVFPDSDELPGAEPDYLNAGKSVRELYETASPNIYEGKYTPVVLWDKKLKTVVNNESSEI 148

Query: 114 LRLFTTDFSDLHQEDAPVLFQSQETASLIDNDIKDINKNFQSLMYKLAFLDKQADYDTYSK 173
 +R+F T+F+ + + L+ +I+ + + +YK F KQ Y+
30 Sbjct: 149 IRMFNTEFNIGIAKTPSLDLYPSHLRDVINETNGWVFENGINNGVYKCGFARKQEPYNEAVN 208

Query: 174 EFFTFLDQKEHLGQRPFLGDNLSEVDIHFFTPLVRWDIAGRDLNQLKALEDYDPNIF 233
 + + +D+ E +LG++ ++ G+ +E DI F L+R+D N++ L +YPNIF
Sbjct: 209 QLYEAVDRCEVVLGKQRYICGNFTFEADIRLFVTLIRFDEVYAVHFKCNKRLLREYPNIF 268

35 Query: 234 SWAKTLYNDFNLKTLTNPQSIKNNYY-----LGKFGRAVRHHTIVPTGPNM 279
 ++ K +Y + + N + IK +YY + FG I+P GPN+
Sbjct: 269 NYIKDIYQIHGMSSTVNMEHIKQHYYGSHPTINPFG-----IIPHGPNI 312

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1928

A DNA sequence (GBSx2037) was identified in *S.agalactiae* <SEQ ID 5975> which encodes the amino acid sequence <SEQ ID 5976>. Analysis of this protein sequence reveals the following:

45 Possible site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
50 bacterial cytoplasm --- Certainty=0.2647(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:BAB07793 GB:AB037666 hypothetical protein [Streptomyces sp.
 CL190]

-2171-

Identities = 127/331 (38%), Positives = 194/331 (58%), Gaps = 9/331 (2%)

Query: 4 RKDDHIKYALKYQSHY---NSFDDIELIHSSSLPKYNVNDIDLSTHFAGQSFEFFPYINAM 60
 5 RKDDH++ A++ + + N FDD+ +H +L + D+ L+T FAG S++ P YINAM
 Sbjct: 6 RKDDHVRLAIEQHNASHGRNQFDDVSFVHHALAGIDRPDVSLSATSFAGISWQVPIYINAM 65

Query: 61 TGGSEKKGKAVNHKLAQVAQATGIVMVTGSYSAALKNDE--DDSYPTDLYPDLKLATNIG 118
 TGGSEK +N LA A+ TG+ + +GS +A +K+ D D P+ + NI
 10 Sbjct: 66 TGGSEKTGLINRDLATAARETGVPPIASGSMNAYIKDPSCADTFRVLRDENPNNGFVIANIN 125

Query: 119 LDKPVPAAESTVKAMNPIFLQVHVNMVQELLMPEGEREFHMRSHLKEYVDNIQCPLILK 178
 V A+ + + LQ+H+N QE MPEGR F W +++ + P+I+K
 Sbjct: 126 ATTTVDNAQRAIDLIEANALQIHINTAQETPMPEGDRSFASWVPQIEKIAAAVDIPVIVK 185

Query: 179 EVGFGMDLQSIKDAYDIGITTVDISGRGGTSFAYIENQRGR--DRSYLNTWGQTTAQSLI 236
 EVG G+ Q+I D+G+ D+SGRGGT FA IEN R D ++L+ WGQ+TA L+
 15 Sbjct: 186 EVGNGLSRQTILLADLVQVQADVSGRGGTDFARIENGRRELGDYAFHLHGWGQSTAACL 245

Query: 237 NAQSMMDKMDILASGGIRHPLDMVKCLVLGAKAVGLSRTVLELVERYPVDDVIAILNSWK 296
 +AQ + + +LASGG+RHPLD+V+ L LGA+AVG S L + VD +I L +W
 20 Sbjct: 246 DAQDI--SLPVLASGGVRHPLDVRALALGARAVGSSAGFLRTLMDGDVDAITKLTW 303

Query: 297 EDLRMIMCALNCKKITDLRQVNYILYQQLKE 327
 + L + L + DL + + +L+G+L++
 25 Sbjct: 304 DQLAALQTMIGARTPADLTRCDVLLHGEIRD 334

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5977> which encodes the amino acid sequence <SEQ ID 5978>. Analysis of this protein sequence reveals the following:

Possible site: 51
 30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2823(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 244/329 (74%), Positives = 284/329 (86%)

Query: 1 MTNRKDDHIKYALKYQSHYNSFDDIELIHSSSLPKYNVNDIDLSTHFAGQSFEFFPYINAM 60
 40 MTNRKDDHIKYALKYQS YN+FDDIELIH SLP Y+++DIDLSTHFAGQ F+FFPYINAM
 Sbjct: 31 MTNRKDDHIKYALKYQSPYNADFDDIELIHSLPSYDLSIDLSTHFAGQDFDFPYINAM 90

Query: 61 TGGSEKKGKAVNHKLAQVAQATGIVMVTGSYSAALKNDEDDSYPTDLYPDLKLATNIGLD 120
 45 TGGG+KGKAVN KLA+VA ATGIVMVTGSYSAALKN DDSY ++ +LKLATNIGLD
 Sbjct: 91 TGGSQKKGKAVNEKLAKVAAATGIVMVTGSYSAALKNPNDDSYRLHEVADNLKLATNIGLD 150

Query: 121 KPVPAAEESTVKAMNPIFLQVHVNMVQELLMPEGEREFHMRSHLKEYVDNIQCPLILKEV 180
 KPV + TV+ M P+FLQVHVNMVQELLMPEGER FH W+ HL EY I P+ILKEV
 50 Sbjct: 151 KPVALGQQTVQEMQPLFLQVHVNMVQELLMPEGERVFTWKKHLAEYASQIPVPVILKEV 210

Query: 181 GFGMDLQSIKDAYDIGITTVDISGRGGTSFAYIENQRGRDRSYLNTWGQTTAQSLINAQS 240
 GFGMD+ SIK A+D+GI T DISGRGGTSFAYIENQRG DRSYLN WGQTT Q L+NAQ
 55 Sbjct: 211 GFGMDVNSIKLAHDLGIQTFDISGRGGTSFAYIENQRGGDRSYLNDWGQTTVQCLLNAQ 270

Query: 241 MMDKMDILASGGIRHPLDMVKCLVLGAKAVGLSRTVLELVERYPVDDVIAILNSWKEDLR 300
 +MD+++ILASGG+RHPLDM+KC VLGA+AVGLSRTVLELVE+YP + VIAI+N WKE+L+
 Sbjct: 271 LMDQVEILASGGVRHPLDMIKCFVLGARAVGLSRTVLELVEKYPTERVIAIVNGWKEBELK 330

Query: 301 MIMCALNCKKITDLRQVNYILYQQLKEAN 329
 +IMCAL+CK I +L+ V+Y+LYG+L++ N
 60 Sbjct: 331 IIMCALDCKTIKELKGVLYLLYGRLLQV 359

-2172-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1929

A DNA sequence (GBSx2038) was identified in *S.agalactiae* <SEQ ID 5979> which encodes the amino acid sequence <SEQ ID 5980>. This protein is predicted to be phosphomevalonate kinase. Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0785(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG02457 GB:AF290099 phosphomevalonate kinase [Streptococcus pneumoniae]
Identities = 170/330 (51%), Positives = 233/330 (70%), Gaps = 1/330 (0%)

Query: 1 MVKVQTGGKLYIAGEYAILYPGQVAILKNVPIYMTALATFADNYSLSYSDMFNYTASLQPD 60
M+ V+T GKLY AGEYAIL PGQ+A++K++PIYM A F+D+Y +YSDFM++ L+P+
Sbjct: 1 MIAVKTGKLYWAGEYAILLEPGQLALIKDIPYMRATIAFSDSYRIYSDFMDFAVDLRPN 60

Query: 61 KQYSLIQETILLMEEWLINFGNKIKPIHLEITGKLERYGLKFGIGSSGSVVVLTIKAMAA 120
YSLIQETI LM ++L G+N++P L+I GK+ER G KFG+GSSGSVVVL +KA+ A
Sbjct: 61 PDYSLIQETIALMGDFLAVRGQNLRPFSLKICGKMEREGKKFGLGSSGSVVVLVVKALLA 120

Query: 121 LYEIEMPSDLLFKLSAYVLLKRGDNGSMGDIACIAYEHLISYSAFDRRAVSKMIETKPLE 180
LY + + +LLFKL++ VLLKRGDNGSMGD+ACI E L+ Y +FDR+ + +E + L
Sbjct: 121 LYNLSDVQNLFLKLTSAVLLKRGDNGSMGDLACIADLVLYQSPDRQKAAWLEENLA 180

Query: 181 QVLEAEWGYRITKIQALLEMDFLVGWTMQPSISKEMINIVKSTITQRFLLDTKYQVVQLL 240
VLE +WG+ I++++ LE DFLVGWT + ++S M+ +K I Q FL +K VV L+
Sbjct: 181 TVLERDWGFFISQVKPTLECDFLVGWTKAVVSSHVMVQIKQINQNLSSSKETVVSLV 240

Query: 241 SAFKEGDKEAIKRCLEELISLLLFNLHPSIYTDKLQKLKEASKGLDIVTKSSGSGGDCGI 300
A ++G E + +E S LL L IYT L++LKEAS+ L V KSSG+GGGDCGI
Sbjct: 241 EALEQCKAEKVIEQVEVASKLLEGLSTDIYTPLLRQLKEASQDLQAVAKSSGAGGDCGI 300

Query: 301 AISFN-KNDNQTLIKRWESAGIELLSKETL 329
A+SF+ ++ TL RW GIELL +E +
Sbjct: 301 ALSFDAQSSRNTLKNRWADLGIELLYQERI 330

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5981> which encodes the amino acid sequence <SEQ ID 5982>. Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2669(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 171/325 (52%), Positives = 227/325 (69%), Gaps = 2/325 (0%)

Query: 4 VQTGGKLYIAGEYAILYPGQVAILKNVPIYMTALATFADNYSLSYSDMFNYTASLQPDKQY 63
VQTGGKLY+ GEYAIL PGQ A++ +P+ MTA + A + L SDFM++ A + PD Y
Sbjct: 22 VQTGGKLYLTGEYAILTPGQKALIHFIPLMMTAEISPAHIQLASDMFSGKAGMTPDASY 81

-2173-

Query: 64 SLIQETILLMEEWLINFGNKINPIHLEITGKLERYGLKFGIGSSGSVVLTIKAMAALYE 123
 +LIQ T+ ++L ++P L ITGK+ER G KFGIGSSGSV +LT+KA++A Y+
 Sbjct: 82 ALIQATVKTTFADYLGQSIDQLEPPSLIITGKMERDGGKFGIGSSGSVTLTLKALSAYYQ 141

Query: 124 IEMPSDLLFKLSAYVLLKRGDNGSMGDIACIAYEHLISYSAFDRRAVSKMIETKPLEQVL 183
 I + +LLFKL+AY LLK+GDNGSMGDIACIAY+ L++Y++FDR VS ++T PL+++L
 Sbjct: 142 ITLTPPELLFKLAAYTLLKQGDNGSMGDIACIAYQTLVAYTSFDREQVSNWLQTMPLKLL 201

Query: 184 EAEWGYRITKIQALLEMDFLVGWIMQPSISKEMINIVKSTITQRFLDDTKYQVVQ-LLSA 242
 +WGY I IQ L DFLVGWT P+IS++MI V ++IT FL T YQ+ Q + A
 Sbjct: 202 VKDWGYHIQVIQPALPCDFLVGWTKIPAI SRQMIQQVTASITPAFL-RTSYQLTQSAMVA 260

Query: 243 FKEGDKAEIKRCLEBISLLFNHPSIYTDKLQKLKEASKGLDIVTKSSGSGGDCGIAI 302
 +EG KE +K+ L S LL LHP+IY KL L A + D V KSSGSGGDCGIA+
 Sbjct: 261 LQEGHKEELKKS LAGASHLLKELHPAIYHPKLVTLVAACQKQDAVAKSSGSGGDCGIAL 320

Query: 303 SFNKNNDQTLIKRWESAGIELLSKE 327
 +FN++ TLI +W+ A I LL +E
 Sbjct: 321 AFNQDARDTLISKWQEADIALLYQE 345

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1930

A DNA sequence (GBSx2039) was identified in *S.agalactiae* <SEQ ID 5983> which encodes the amino acid sequence <SEQ ID 5984>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -1.75 Transmembrane 20 - 36 (18 - 36)

----- Final Results -----
 bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1931

A DNA sequence (GBSx2040) was identified in *S.agalactiae* <SEQ ID 5985> which encodes the amino acid sequence <SEQ ID 5986>. This protein is predicted to be mevalonate diphosphate decarboxylase. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1557(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG02456 GB:AF290099 mevalonate diphosphate decarboxylase
 [Streptococcus pneumoniae]

-2174-

Identities = 219/312 (70%); Positives = 264/312 (84%)

Query: 1 MDGKSISVKS YANIAI I KYWGKADAEK MIPATSSISLTLENMYTETRLTALGKDAKKDEF 60
 MD + ++V+SYANIAI I KYWGK ++M+PATSSISLTLENMYTET L+ L + DEF
 5 Sbjct: 1 MDREPVTVR SYANIAI I KYWGKKKEKEMVPATSSISLTLENMYTETTLSP L PANVTADDEF 60

Query: 61 YISGVLQNDHEHDKMSAILDRFRQNRSGFVKIETTNNMPTAAGLSSSSSGLSALVKACND 120
 YI+G LQN+ EH KMS I+DR+R GFV+I+T NNMPTAAGLSSSSSGLSALVKACN
 10 Sbjct: 61 YINGQLQNEVEHAKMSKIIDRYRPAGEGFVRIDTQNNMPTAAGLSSSSSGLSALVKACNA 120

Query: 121 FFGTNLSQS QLAQEAKFASGSSSR SFFGPVAAWDKDSGDIYKVHTNLDLAMI MLVLNDKR 180
 +F L +S QLAQEAKFASGSSSR SF+GP+ AWDKDSG+IY V T+L LAMI MLVL DK+
 Sbjct: 121 YFKGLGDRS QLAQEAKFASGSSSR SFYGPLGAWDKDSGEIYPVETDLKLAMI MLVLEDKK 180

Query: 181 KPISSREGMKICTETSTTFNEWVRQSEQDYQDMLVYLKNND FQKVGQLTERNALAMHSTT 240
 KPISSR+GMK+C ETSTTF++WVRQSE+DYQDML+Y LK NDF K+G+LTE+NALAMH+TT
 15 Sbjct: 181 KPISSRDGMKLCVETSTTFDQWVRQSEKDYQDMLIYLKENDFAKIGELTEKNALAMHATT 240

Query: 241 KTATPAFSYLT EET YKAMDVVKKLREKGHECYTMDAGPNVKVLCRLQDLEALAAILEKD 300
 KTA+PAFSYLT+ +Y+AM V++LREKG CY+TMDAGPNVKV C +DLE L+ I +
 20 Sbjct: 241 KTASPAFSYLT DASYEAMAFVRQLREKGEACYFTMDAGPNVKVFCQEKDLEHLSEIFGQR 300

Query: 301 YRIIVSTTKELA 312
 YR+IVS TK+L+
 25 Sbjct: 301 YRLIVSKTKDLS 312

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5987> which encodes the amino acid sequence <SEQ ID 5988>. Analysis of this protein sequence reveals the following:

Possible site: 36
 30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1271(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 221/313 (70%), Positives = 258/313 (81%)

Query: 1 MDGKSISVKS YANIAI I KYWGKADAEK MIPATSSISLTLENMYTETRLTALGKDAKKDEF 60
 +D I+V SYANIAI I KYWGK + KMIP+TSSISLTLENM+T T ++ L A D+F
 40 Sbjct: 1 VDPNVITVTSYANIAI I KYWGKENQAKMIPSTSSISLTLENMFTTTSVSFLPDTATSDQF 60

Query: 61 YISGVLQNDHEHDKMSAILDRFRQNRSGFVKIETTNNMPTAAGLSSSSSGLSALVKACND 120
 YI+G+LQND EH K+SAI+D+FRQ FVK+ET NNMPTAAGLSSSSSGLSALVKAC+
 45 Sbjct: 61 YINGILQND EETKISAIIDQFRQPGQAFVKMETQNNMPTAAGLSSSSSGLSALVKACDQ 120

Query: 121 FFGTNLSQS QLAQEAKFASGSSSR SFFGPVAAWDKDSGDIYKVHTNLDLAMI MLVLNDKR 180
 F T L Q LAQ+AKFASGSSSR SFFGPVAAWDKDSG IYKV T+L +AMIMLV LN +
 50 Sbjct: 121 LFD TQLDQKALAQKAKFASGSSSR SFFGPVAAWDKDSGAIYKVETDLKMAMI MLVLNAK 180

Query: 181 KPISSREGMKICTETSTTFNEWVRQSEQDYQDMLVYLKNND FQKVGQLTERNALAMHSTT 240
 KPISSREGMK+C +TSTTF++WV QS DYQ ML YLK N+F+KVGQLTE NALAMH+TT
 55 Sbjct: 181 KPISSREGMKLCRDTSTTFDQWVEQSAIDYQHMLTYLKTNNFEKVGQLTEANALAMHATT 240

Query: 241 KTATPAFSYLT EET YKAMDVVKKLREKGHECYTMDAGPNVKVLCRLQDLEALAAILEKD 300
 KTA P FSYLT+E+Y+AM+ VK+LR++G CY+TMDAGPNVKVLC L +DL LA L K+
 60 Sbjct: 241 KTANPPFSYLT KESYQAMEAVKELRQEGFACYFTMDAGPNVKVLCLEKDLAQLAERLGKN 300

Query: 301 YRIIVSTTKELAD 313
 YRIIVS TK+L D
 Sbjct: 301 YRIIVSKTKDLDP 313

-2175-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1932

5 A DNA sequence (GBSx2041) was identified in *S.agalactiae* <SEQ ID 5989> which encodes the amino acid sequence <SEQ ID 5990>. Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1512(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5991> which encodes the amino acid sequence <SEQ ID 5992>. Analysis of this protein sequence reveals the following:

Possible site: 35
>>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1117(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

25 Identities = 182/290 (62%), Positives = 223/290 (76%)

Query: 1 MKEKFGIGKAHSKIILMGEHSVVYGYPAIAIPLKNIEVTCLIEEAPQLIALDMDPLSTA 60
M E G GKAHSKIIL+GEH+VVYGYPAIA+PL +IEV C I A + + D D LSTA

30 Sbjct: 6 MNENIGYGKAHSKIILIGEHAVVYGYPAIALPLTDIEVVCHIFPADKPLVDFDYDTLSTA 65

Query: 61 IFAALDYLKGTSSKIAYHIESQVPERGMGSSAAVAIAAIRAVFDYFDELDLEADLLECLV 120
I+A+LDYL + IAY I SQVP++RGMGSSAAV+IAAIRAVF Y E L DLLE LV

Sbjct: 66 IYASLDYLQRLQEPYAEIVSQVPQKRGMGSSAAVSAIAAIRAVFSYCCQEPLSDDLLEILV 125

35 Query: 121 NRAEMIAHSNPSGLDAKTCLSENTIKFIRNIGFSTVPMHLNAYLVIADTGIHGHTKEAVD 180
N+AE+IAH+NPSGLDAKTCLS++ IKFIRNIGF T+ + LN YL+IADTGIHGHT+EAV+

Sbjct: 126 NKAETIAHTNPSGLDAKTCLSDHAIKFTIRNIGFETIBIALNGYLIADTGIHGHTREAVN 185

40 Query: 181 KVKSSGEAVLPFLKELGYLAASEDAIHKSDSKQLGSLMTKAHQSLKQLGVSSLEADHLV 240
KV E LP+L +LG L +A E AI++ + +G LMT+AH +LK +GVS +AD LV

Sbjct: 186 KVAQFEETNLFYLAKLGALTQALERAINQKNKVAIGQLMTQAHSALKAGVSISKADQLV 245

Query: 241 EVAISCGALGARMSGGGLGGCIIALVKEKREAERLSQQLEREGAVNTWTE 290
E A+ GALGAKM+GGGLGGC+IAL K AE++S +L+ EGAVNTW +

45 Sbjct: 246 EAALRAGALGAKMTGGGLGGCMIALADTKDMAEKISHRLKEEGAVNTWIQ 295

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1933

50 A DNA sequence (GBSx2042) was identified in *S.agalactiae* <SEQ ID 5993> which encodes the amino acid sequence <SEQ ID 5994>. This protein is predicted to be a histidine protein kinase. Analysis of this protein sequence reveals the following:

Possible site: 26
>>> Seems to have an uncleavable N-term signal seq

55 INTEGRAL Likelihood = -13.43 Transmembrane 12 - 28 (4 - 33)

-2176-

INTEGRAL Likelihood = -9.29 Transmembrane 163 - 179 (157 - 191)

----- Final Results -----

5 bacterial membrane --- Certainty=0.6371(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAF79919 GB:AF039082 putative histidine protein kinase
[Lactococcus lactis]
Identities = 78/315 (24%), Positives = 154/315 (48%), Gaps = 33/315 (10%)

Query: 101 SDRQIKNYAKRIVSQNSHSGHITYNFSTYSYLLKKVGNKYLVVFLDTTNQYLDNQRLQ 160
+++QI N + + +N + + Y + T S + V++ + Q +

15 Sbjct: 84 NEKQI-NTIQTVSVKNPYGDNWHYRYLTTSQFIITNSDGTVPVYVQIFSNVDQIQDAMS 142

Query: 161 LSIWM---SLVSFIVFMVIVSV-LSGRVILPFVANYEKQRRFITNAGHELKTPLAIISAN 216
++W+ ++++F + VI+S+ L+ + P +A YEKQ+ F+ NA HEL+TPLAI+

20 Sbjct: 143 RAMWVIVTMITFWILSVIISLYLANWTLKPILAAAYEKQKEFVENASHELRTPLAILQNR 202

Query: 217 NELV-----EMMSGSEWTKSTNDQIQRLTGLINGMVSLAR-----FEEQPDISM---- 261
EL+ + +SE + +++ + L + +++LAR E +P +

Sbjct: 203 LELLFQKPTATIIDQSENISESLSEVRNMRLTSLNLLNLARRDSGIKIEPEPTTATYFEN 262

25 Query: 262 VDLDFSHITKDAEDFKGPIIKDGKDFIMSIQPGIHVKAEEKSLFELVTLVDNANKYCD 321
+ + +T++A + F G + +G V ++ + +L+T+L DNA KY D

Sbjct: 263 IFNSYEMLTENAGKKFSGNLKLEGT-----VNLDQALIKQLLTILFDNALKYTD 311

30 Query: 322 PMGTIVTVKLSRSSRLRRAKLEVSNTYKNGKDIDYSKFFERFYREDESHNNKSGYGIGLS 381
G ++V + ++ V++ + D D K F+RF+R D++ +K G G+GLS

Sbjct: 312 SEGEISVDVIKNGGF--LTFVADNNGEGISDEDKKKIFDRFRVDKARTRQKGGGLGLGLS 369

Query: 382 IVTSLVHLFKGSIDV 396
+ +V + G I V

35 Sbjct: 370 LAKQIVEAYNGKITV 384

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5751> which encodes the amino acid sequence <SEQ ID 5752>. Analysis of this protein sequence reveals the following:

40 Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood =-11.30 Transmembrane 18 - 34 (13 - 42)
INTEGRAL Likelihood =-10.35 Transmembrane 170 - 186 (163 - 199)

45 ----- Final Results -----
bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

50 Identities = 233/410 (56%), Positives = 303/410 (73%), Gaps = 1/410 (0%)

Query: 1 MFRNRLRLRFIGIAALAILVVLFSVGVLSANHYQTKNEIYRVLTILADNNGRIPNKLEF 60
MF +R+RFI IA++AI ++L S+VG++N+A YQ++ EI R+L +++ N G++P E

55 Sbjct: 10 MFRNIRIRFIMIASIAIFILSSIVGIINTARCYQSQQEINRILHLISSNKGKLPGTTES 69

Query: 61 SKELGDDLSTDAIFQFRYFSARTDAKGNVTSFDSRNIFEVSQRIKRYAKRIVSQNSHSG 120
SK LG LS D++ QFRY+S +A G++ S ++ NI + + + +A+ G

60 Sbjct: 70 SKRLGTLKLSLSDSLSQFRYYSVIFNANGHLLSSNTANISALDREEAQYFARLFAKSGEKG 129

Query: 121 HITYNFSTYSYLLKKVGNKYLVVFLDTTNQYLDNQRLQLSIWMSLVSFIVFMVIVSVL 180
+ S YSYL+ ++ + LVV LD TT + LL +S+ ++ FI F+V+VS+

Sbjct: 130 SYRHQDSVYSYLITQLPNEEKLVVILDTTTFYFRSVGDLAVSVMLAFGGFIFVVLVLSLF 189

Query: 181 SGRVILPFVANYEKQRRFITNAGHELKTPLAIISANNELVEMMSGSEWTKSTNDQIQRL 240

-2177-

SG VI PFV NYEKQRRFITNAGHELKTPLAIISSANNELVE+M+GESEWTKST+DQ++RL
 Sbjct: 190 SGMVVKPFVQNYEKQRRFITNAGHELKTPLAIISSANNELVELMTGESEWTKSTSDQVKRL 249

Query: 241 TGLINGMVSLARFEEQPDISMVDLDFSHITKDAEDFKGPIIKDGKDFIMSIQPGIHVKA 300
 TGLIN M++LAR EEQPD+ + +DFS I +DAAEDFK ++KDGK F ++IQP I +KA
 Sbjct: 250 TGLINQMITLARLEEQPDVVLHVMVDFSAIAQDAAEDFKSLVLKDGKRFDLTIQPNIMIKA 309

Query: 301 EEKSLFELVTLVDNANKYCDPMGTVTVKLSRSSRLR-RAKLEVSNTYKNGKDIDYSKFF 359
 EEKSLFELVT+LVDNANKYCDP G V V L+ R R RAKLEVSNTY GK IDYS+FF
 Sbjct: 310 EEKSLFELVTLVDNANKYCDPKGLVKVSLTTIGRRRKRAKLEVSNTYLEGKSIDYSRFF 369

Query: 360 ERFYREDESHNNKSGYGIGLSIVTSLVHLFKGSIDVNYKHDTITFVIYI 409
 ERFYREDESHN+K+ GYGIGLS+ S+V LFKG+I VNYK+D I F + I
 Sbjct: 370 ERFYREDESHNSKEGYGIGLSMAESMVKLFGKTITVNYKNDIAIVFTTVVI 419

SEQ ID 5994 (GBS273) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 14; MW 46kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 5; MW 71kDa).

GBS273-GST was purified as shown in Figure 208, lane 4.

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1934

A DNA sequence (GBSx2043) was identified in *S.agalactiae* <SEQ ID 5995> which encodes the amino acid sequence <SEQ ID 5996>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2181(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1935

A DNA sequence (GBSx2044) was identified in *S.agalactiae* <SEQ ID 5997> which encodes the amino acid sequence <SEQ ID 5998>. This protein is predicted to be two-component response regulator (trcR).

- Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2503(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- A related GBS nucleic acid sequence <SEQ ID 9379> which encodes amino acid sequence <SEQ ID 9380> was also identified.

-2178-

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04091 GB:AP001508 two-component response regulator [Bacillus halodurans]
Identities = 71/183 (38%), Positives = 120/183 (64%), Gaps = 3/183 (1%)

5 Query: 9 RVLIAEDEEQMSRVLSTAISHQGYVVDVAYDQQTALANQAYDVMVMDVMPVKTGIE 68
R+LI EDE+++RVL + H+GY D A+ G ++ +A+D+++DVM+P +G+E
Sbjct: 3 RILIIIEDEKKIARVLQLELEHREGYETDAAFSGSDGLETFQAHAWDLVLLDVMLPELSGLE 62

10 Query: 69 AVKEIRQSGNKSHIIMLTAMAEIDRVGTGLDAGADDYLTKEPFLKELLARLRSMRRL- 127
++ IR + + II+LTA I D+V+GLD GA+DY+TKPF ++ELLAR+R+ R ++
Sbjct: 63 VLRRIRMTDPVTPITILLTARNISIPDKVSGLDLGANDYITKPFIEELLARVRACLRVQT 122

15 Query: 128 -DFTPNVLSLGRVTLVSGEQELQCEN-TIRLAGKEAKMLAFFMLNHDKELSTQQLFEHVW 185
+ + L +T++ +++Q N TI L KE ++L FF+ N + LS +Q+ +VW
Sbjct: 123 RERVEDTLMFQELTINEKTRDVQRNETIELTPKEFELLVFFIKKQVLSREQILTNVW 182

Query: 186 GAD 188
G D
Sbjct: 183 GFD 185

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5999> which encodes the amino acid sequence <SEQ ID 6000>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2391(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 125/185 (67%), Positives = 151/185 (81%)

35 Query: 8 MRVLI AEDEEQMSRVLSTAISHQGYVVDVAYDQQTALANQAYDVMVMDVMPVKTGI 67
M++L+AEDE QMS VL+TA++HQGY VDV ++GQ AID A NAYD+M++D+MMP+K+GI
Sbjct: 1 MKILLAEDEWQMSNVLTATANTHQGYDVDVFNQQAIDKAKONAYDIMILDIMMPIKSGI 60

40 Query: 68 EAVKEIRQSGNKSHIIMLTAMAEIDRVGTGLDAGADDYLTKEPFLKELLARLRSMRRL- 127
EA+KEIR SGN SHIIMLTAMAEI+DRVGTGLDAGADDYLTKEPFLKELLARLRSM RR+E
Sbjct: 61 EALKEIRASGNCNCSIIMLTAMAEINDRVGTGLDAGADDYLTKEPFLKELLARLRSMERVE 120

Query: 128 DFTPNVLSLGRVTLVSGEQELQCEN-TIRLAGKEAKMLAFFMLNHDKELSTQQLFEHVWGA 187
FTP VI. VTT++ EORT. N TRLA KE KLLAF MIN K T. TL TLLHGW

-2179-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:BAB05604 GB:AP001513 unknown conserved protein [Bacillus halodurans]
Identities = 67/182 (36%), Positives = 111/182 (60%), Gaps = 4/182 (2%)

Query: 17 LEDFSQRIQLENDKAKVETGYKLYEHIIGRIKTSDSMIEKCRKQLPVTVD SALKTIRDS 76
L++ + +I + + + Y EH+ R+K+ +S++ K +R+ T++S + +RD

10 Sbjct: 29 LQELNTKIDILKQEFQYIHDYNPIEHVSSRVKSPESIVNKIQRRGNDFTLESIRENVRDI 88

Query: 77 IGVRIICGFVNDIYQIIERIKAFDDCRIVVEKDVIQHVKNPGYRSYHVILEIDTPYPDC 136
G+RI C F +DIY + E++ D +V KDYI++ KPNYRS H+IL I P +

15 Sbjct: 89 AGIRITCSFESDIYTLSEQLMQHQHDSVVETKDYIKNPKNPGYRSLHLILSI----PIFM 144

Query: 137 GNSDGKYYIEIQLRTIAQDSWASLEHQMKYKHDIENPERIVRELKRCADEMASVDLT 196
+ Y+E+Q+RTIA D WASLEH++ YK++ PE +++ELK A+ A +D M+

Sbjct: 145 SDRVQDVYVEVQIRTIAMDFWASLEHKIYYKYNKNVPEHLLKELKDAESAALLDQKMEK 204

20 Query: 197 IR 198
I+

Sbjct: 205 IQ 206

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6003> which encodes the amino acid
25 sequence <SEQ ID 6004>. Analysis of this protein sequence reveals the following:

Possible site: 33
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1057(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 127/206 (61%), Positives = 162/206 (77%)

Query: 3 TNIYGDYGRYLPLILEDFSQRIQLENDKAKVETGYKLYEHIIGRIKTSDSMIEKCRKQL 62
++IY + YLPL+L+ + I EN K+K ETG+KLYEH RIK+ SMIEKC+RKQL

40 Sbjct: 11 SSIYSGFEVYLPLVLQTITDVIIAENIKSKKETGFKLYEHFTSRIKSEASMIEKCQRKQL 70

Query: 63 PVTVD SALKTIRDSIGVRIICGFVNDIYQIIERIKAFDDCRIVVEKDVIQHVKNPGYRSY 122
P+T SALK I+DSIG+RIICGF++DIY++++ +K+ + EKDYI + KPNYRSY

Sbjct: 71 PLTSK SALKI IKDSIGIRIICGFIDDIYRMVDLLKSIPGMSVNTKDYILNAKPNYRSY 130

45 Query: 123 HVILEIDTPYPDC LGNSDGKYYIEIQLRTIAQDSWASLEHQMKYKHDIENPERIVRELKR 182
H+ILE++T +PD LG G Y+IE+QLRTIAQDSWASLEHQMKYKH + N E I RELKR

Sbjct: 131 HLILELETHFFDILGEKKGCYFIEVQLRTIAQDSWASLEHQMKYKHQVANAEMITRELKR 190

50 Query: 183 CADEMASVDLTMQTIRQLIESGTKKE 208
CADE+AS D+TMQTIRQLI+ T++E

Sbjct: 191 CADELASCDVTMQTIRQLIQETTEEE 216

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
vaccines or diagnostics.

55 Example 1937

A DNA sequence (GBSx2046) was identified in *S.agalactiae* <SEQ ID 6005> which encodes the amino
acid sequence <SEQ ID 6006>. Analysis of this protein sequence reveals the following:

Possible site: 40
>>> Seems to have no N-terminal signal sequence

-2180-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3250(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA37193 GB:X53013 ORF1 (AA 1 - 384) [Lactococcus lactis]
 Identities = 30/55 (54%), Positives = 37/55 (66%)

10

Query: 1 MEFYKTLKRKFINDADTIFIEQSQFEIFYIETDHNSSSSHVLDYQSQKEFEK 55
 ME +YKTLKR+ INDA ++ EIF YIET +N+ H LDYQS K+FEK
 Sbjct: 327 MESFYKTLKRELINDAHFETRAEATQEIFKYIETYYNTKWMHSGLDYQSPKDFEK 381

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6007> which encodes the amino acid sequence <SEQ ID 6008>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

20

----- Final Results -----

 bacterial cytoplasm --- Certainty=0.3065(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 31/59 (52%), Positives = 39/59 (65%)

Query: 1 MEFYKTLKRKFINDADTIFIEQSQFEIFYIETDHNSSSSHVLDYQSQKEFEKIITN 59
 ME +YKTLKR+ +NDA I+Q+Q EIF Y ET +N H L Y S EFEKI+T+
 30 Sbjct: 13 MEAFYKTLKRELVNDAHFATIKQAQLEIFKYSETIYYPKRLHSALGYLSPVEFEKIVTH 71

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1938

35 A DNA sequence (GBSx2047) was identified in *S.agalactiae* <SEQ ID 6009> which encodes the amino acid sequence <SEQ ID 6010>. This protein is predicted to be R5 protein. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence
 40 INTEGRAL Likelihood = -3.98 Transmembrane 30 - 46 (29 - 51)
 INTEGRAL Likelihood = -2.76 Transmembrane 967 - 983 (966 - 985)

----- Final Results -----

45 bacterial membrane --- Certainty=0.2593(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8935> which encodes amino acid sequence <SEQ ID 8936> was also identified. Analysis of this protein sequence reveals the following:

50 Lipop: Possible site: -1 Crend: 8
 SRCFLG: 0
 McG: Length of UR: 2
 Peak Value of UR: 2.44
 Net Charge of CR: 2
 55 McG: Discrim Score: 0.78
 GvH: Signal Score (-7.5): -0.0599995

-2181-

Possible site: 39
 >>> Seems to have a cleavable N-term signal seq.
 Amino Acid Composition: calculated from 40
 ALOM program count: 0 value: 7.37 threshold: 0.0
 5 PERIPHERAL Likelihood = 7.37. 194
 modified ALOM score: -1.97
 *** Reasoning Step: 3
 10 Rule gpol
 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 15 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
 LPXTG motif: 944-948

No corresponding DNA sequence was identified in *S.pyogenes*.

20 SEQ ID 8936 (GBS200) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 3; MW 107.4kDa), in Figure 169 (lane 4; MW 122kDa) and in Figure 238 (lane 11; MW 122kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 3; MW 132kDa).

Purified Thio-GBS200-His is shown in Figure 244, lane 9.

25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1939

A DNA sequence (GBSx2048) was identified in *S.agalactiae* <SEQ ID 6011> which encodes the amino acid sequence <SEQ ID 6012>. This protein is predicted to be a 16.1 kDa transcriptional regulator. Analysis
 30 of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.3919(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9953> which encodes amino acid sequence <SEQ ID 9954>
 40 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB16108 GB:Z99124 similar to transcriptional regulator (MarR family) [Bacillus subtilis]
 Identities = 30/114 (26%), Positives = 59/114 (51%), Gaps = 3/114 (2%)
 45 Query: 29 DVEHLAGPQGHVLMYLYKHPDKMSIKAVEEILHISKSVASNLVKRMEKNGFIAIVPSKT 88
 D++ G +LV +Y++P + + + E++ + ++ A+ +K++E GFI +P +
 Sbjct: 25 DLDLTRGQYLYLVR-IYENPG--IIQEKLAEMIKVDRTTAARAIAKKLEMCGFIQKLPDEQ 81
 50 Query: 89 DKRVKYLVLTHLGKKKATQFEIFLEKLHSTMLAGITKEEIRTTKKVIRTLAKNM 142
 +K++K L+ T GKK E L+G T EE T ++ + KN+
 Sbjct: 82 NKKIKKLFPTTEKGKKVYPLLRREGEHSTEVALSGFTSBEKETISALLHRVRKNI 135

-2182-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6013> which encodes the amino acid sequence <SEQ ID 6014>. Analysis of this protein sequence reveals the following:

Possible site: 43
>>> Seems to have no N-terminal signal sequence

5 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4175 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

10 An alignment of the GAS and GBS proteins is shown below.

Identities = 27/64 (42%), Positives = 46/64 (71%)

15 Query: 3 MENPLQKARILVNQLEKYLDHYAKEYDVEHLAGPQGHLMVLYKHPDKDMSIKAVEEILH 62
M + R L++Q+E+ D AK+YDVEHLAGPQG++++L KH +++++ +K +E+ L
Sbjct: 1 MSQVIGDLRELHIEQISDEIAKKYDVEHLAGPQGYVLVFLAKHQNEIFVKDIEKQLR 60

Query: 63 ISKS 66
I +S
20 Sbjct: 61 IFQS 64

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1940

25 A DNA sequence (GBSx2049) was identified in *S.agalactiae* <SEQ ID 6015> which encodes the amino acid sequence <SEQ ID 6016>. This protein is predicted to be 5'-nucleotidase family protein. Analysis of this protein sequence reveals the following:

Possible site: 27
>>> Seems to have a cleavable N-term signal seq.

30 INTEGRAL Likelihood = -2.66 Transmembrane 668 - 684 (665 - 684)

----- Final Results -----
bacterial membrane --- Certainty=0.2062 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
35 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12747 GB:Z99108 similar to 5'-nucleotidase [Bacillus subtilis]
Identities = 178/535 (33%), Positives = 270/535 (50%), Gaps = 55/535 (10%)

40 Query: 28 DQVGQVIGVNDFFHGALDNTGTANMPDGKVANAGTAAQLD---AYMDDAQKDFKQINPNG 84
+ V +++++ +ND HG +D ++ DG GT ++D AY+ + + + K
Sbjct: 586 EHVPLRILSMNDLHGKIDQQYELDL-DGNGTVDGTFGRMDYAAAYLKEKKAEEKKN----- 639

45 Query: 85 ESIRVQAGDMVGASPA NSGLLQDEPTVKNFNAMNVEYGT LGNHEFDEGLAEYNRI VTGKA 144
S+ V AGDM+G S S LLQDEPTV+ + + GT+GNHEFDEG E RI+ G
Sbjct: 640 -SLIVHAGDMIGGSSPVSSLLQDEPTVELMEDIGFDVGT VGNHEFDEGTD ELLRLILNG-G 697

50 Query: 145 PAPDSNINNITKSY PHEAAKQEIVVANVIDKVNKQIPYNWKPYAIKNIPVNNKSVNVGFI 204
P +++P +V AN ++ +P+ +N + V V FI
Sbjct: 698 DHPKGTSGYDGNFP-----LVCANC-----KMKSTGEPFLPAYDIINVEGVPVAFI 744

55 Query: 205 GIVTKDIPNLVRKNYEQYEFLEAETIVKYAKELQAKNVKAIIVVLAHVPATSKNDIAEG 264
G+VT+ +V+ + + EF DEA + K A+EL+ K VKAI VLAH+ A + G
Sbjct: 745 GVVTSQAAGMVMPEGIKNIEFTDEATAVNKAAEELKKKGVKAIIVLAHMSAEQNGNAITG 804

Query: 265 EAAEMMKVNQLFPENSVDIVFAGHNHQTNGLVGKTRIVQALSQ GKAYADVRGVLD TDT 324
E+A++ K ++ +D++FA HNHQ NG V IVQA GKA V +D T

-2183-

Sbjct: 805 ESADLANKT-----DSEIDVIFAAHNQVNVNGEVLGKLVQAFEYGAIGVVDVEIDKTT 859

Query: 325 QDFIETPSAKVIAVAPGKKTGSADIQAIVDQANTIVKQVTEAKIGTAEVSVMITRSVDQD 384
+D ++ SA+++ V K AI+ + TI + + +G A V + S D D

5 Sbjct: 860 KDIVK-KSAEIVYVDQSKIBPDVSASAILKKYETIAEPIISEVVGEEAADMEGGYSNDGD 918

Query: 385 NVSPVGLITEAQLAIARKSWPDIDFAMTNNGGIRADLLIKPDGTITWGAAQAVQPFNGI 444
+P+G+LI + A + DFA+ N GGIR L G ITWG +QPFNG+

10 Sbjct: 919 --TPLGNLIADGMRAAMK-----TDFALMNGGGIREAL---KKGPIITWGDLYNIQPFNGV 968

Query: 445 LQVVEITGRDLYKALNEQYDQKQNFLLQIAGLRYTYTDNKEGGEETPFKVVVKAYKSNTEE 504
L +EI G+DL + +N Q I+G +TYT +KE G+ K+ ++G E

15 Sbjct: 969 LTKLEIKGKDLREIINAQISPVFGPDYSISG--FTYTWDKETGKAVDMKM-----ADGTE 1021

Query: 505 INPDAKYKLVINDFLFGGGDGFASFRNAKLLGAINP-----DTEVFMAITDLEK 554
I PDA Y L +N+F+ A ++ LLG NP D E + Y+ ++

Sbjct: 1022 IQPDATYTLTVNNFMATATG--AKYQPIGLLGK-NPVTGPEDLEATVEYVKSFDE 1073

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1607> which encodes the amino acid sequence <SEQ ID 1608>. Analysis of this protein sequence reveals the following:

Possible site: 40
>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -4.67	Transmembrane	662 - 678 (661 - 679)
INTEGRAL	Likelihood = -2.02	Transmembrane	19 - 35 (18 - 35)

25 ----- Final Results -----

bacterial membrane	---	Certainty=0.2869(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

30

An alignment of the GAS and GBS proteins is shown below.

Identities = 415/688 (60%), Positives = 517/688 (74%), Gaps = 21/688 (3%)

35 Query: 1 MKKKIILKSSVLGLVAGTSMFSSVFADQGVGVQVIGVNDPHGALDNTGTANMPDGKVANA 60
MKK ILKSSVL ++ +++ + V ADQV VQ +GVNDFHGALDNTGTA P GK+ NA

Sbjct: 14 MKKYFILKSSVLSILTSFTLLTDVQADQVDVQFLGVNDFHGALDNTGTAYTPSGKIPNA 73

Query: 61 GTAAQLDAYMDDAQKDFKQTNPNGESIRVQAGDMVGASPANSGLLQDEPTVKNFNAMNVE 120
GTAAQL AYMDDA+ DFKQ N +G SIRVQAGDMVGASPANS LLQDEPTVK FN M E

40 Sbjct: 74 GTAAQLGAYMDDAEIDFKQANQDGTSSIRVQAGDMVGASPANSALLQDEPTVKVFNKMKFE 133

Query: 121 YGTILGNHEFDEGLAEYNRIVTGKAPAPDSNINNITKSYPEAAKQEIIVVANVIDKVNKQI 180
YGTILGNHEFDEGL E+NRI+TG+AP P+S IN+ITK Y HEA+ Q IV+ANVIDK K I

45 Sbjct: 134 YGTILGNHEFDEGLDEFNRIMTGQAPDPESTINDITKQYEHASHQTIVIANVIDKKTDKDI 193

Query: 181 PYNWKPYAIKNIPVNNKSVNVGFIGIVTKDIPNLVLRKNYEQYEFLEAETIVKYAKELQ 240
PY WKPYAIK+I +N+K V +GFIG+VT +IPNLVL++NYE Y+FLD AETI KYAKELQ

Sbjct: 194 PYGWKPYAIKDIAINDKIVKIGFIGVVTTEIPNLVLKQNYEHYQFLDVAETIAKYAKELQ 253

50 Query: 241 AKNVKAIIVLAHVPAATSKNDIAEGEAAEMMKVNLQFPENSVDIVFAGHNHQYTNGLVGK 300
++V AIVVLAHVPAATSK+ + + E A +M+KVNQ++PE+S+DI+FAGHNHQYTNG +GK

Sbjct: 254 EQHVAIIVLAHVPAATSKDGVVDHEMATVMEKVNQIYPEHSIDII FAGHNHQYTNGTIGK 313

55 Query: 301 TRIVQALSQKAYADVRGVLDTDTQDFIETPSAKVIAVAPGKKTGSADIQAIVDQANTIV 360
TRIVQALSQKAYADVRG LDTDT DFI+TPSA V+AVAPG KT ++DI+AI++ AN IV

Sbjct: 314 TRIVQALSQKAYADVRGTLDTDTNDFIKTPSANVAVAPGIKTENSIDIKAIINHANDIV 373

Query: 361 KQVTEAKIGTAEVSVMITRSVDQDNVSPVGLITEAQLAIARKSWPDIDFAMTNNGGIRA 420
K VTE KIGTA S I+++ + D SPVG+L T AQL IA+K++P +DFAMTNNGGIR+

60 Sbjct: 374 KTVTERKIGTAINSTISKTENIDKESPVGNLATTAGLTIAKKTFFPTVDFAMTNNGGIRS 433

Query: 421 DLLIKPDGTITWGAAQAVQPFNGILQVVEITGRDLYKALNEQYDQKQNFLLQIAGLRYTY 480
DL++K D TITWGAAQAVQPFNGILQV+++TG+ +Y LN+QYD+ Q +FLQ++GL YTY

65 Sbjct: 434 DLVVKNDRITITWGAAQAVQPFNGILQVIQMTGQHIYDVNLQCYDENQTYFLQMSGLTYTY 493

-2184-

Query: 481 TDNKEGGEETPPFKVVKAYKSNGEINPDACYKLVINDFLFGGDFGASFNAKLLGAINP 540
 TDN +TPFK+VK YK NGBEIN Y +V+NDFL+GGGDFG++F+ AKL+GAIN
 Sbjct: 494 TDNDPKNSDTPFKIVKVYKDNGBEINLTTTYTVVVNDFLYGGGDFGSAFKKAKLIGAIN 553

5 Query: 541 DTEVFMAIYITDLEKAGKKVSVNNKPKIYVTMKMVNETITQNDGTHSIKKLYLDRQNI 600
 DTE F+ YIT+LE +GK V+ K YVT + + T + G HSII K++ +R GN
 Sbjct: 554 DTEAFITYITNLEASGKTVNATIKGVKNYVTSNLESSTKVNSAGKHSIISKVFRNRDNT 613

10 Query: 601 VAQEIVSDTLNQTKSKSTKINPVTIHKQLHQFTAINPMRNYGKPSNSTTVKSKQLPKT 660
 V+ E++SD L T++ + + T +N T+ S LP T
 Sbjct: 614 VSSEVISDLLTSTENTNNSLGKKET-----TTNKNTISSSTLPIT 653

Query: 661 NSEYQGSFLMSVFG-VGLIGIALNTKKK 687
 Y S +M++ + L G+ KK+
 15 Sbjct: 654 GDNKMSPIMTILALISLGGLNAFIKKR 681

SEQ ID 6016 (GBS328) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 4; MW 73kDa). The GBS328-His fusion product was purified (Figure 213, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 268), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1941

A DNA sequence (GBSx2050) was identified in *S.agalactiae* <SEQ ID 6017> which encodes the amino acid sequence <SEQ ID 6018>. This protein is predicted to be peptide deformylase (def-2). Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.70 Transmembrane 55 - 71 (55 - 74)

30 ----- Final Results -----
 bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB09662 GB:Z96934 peptide deformylase [Clostridium
 beijerinckii]
 Identities = 71/136 (52%), Positives = 96/136 (70%)

40 Query: 1 MIKPIVRDTFFLQKSQMASRADVSLAKDLQETLHANQNYCVGMAANMIGSLKRVIINV 60
 MIKPIV+D FL QKS+ A++ D+ + DL +TL AN +CVG+AAANMIG KR+++ V
 Sbjct: 1 MIKPIVKDILFLGQKSEETKNDMVVIDDLIDTLRANLEHCVGLAANMIGVKKRILVFTV 60

45 Query: 61 GITNLVFMFNPVVVAKSDPYETEESCLSLVGCSTORYCHITISYRDINWKEQQIKLTDFP 120
 G + M NPV++ K PYETEESCLSL+G R T+RY I ++Y D N+ +++ F
 Sbjct: 61 GNLIVPMINPVILKKEKPYETEESCLSLIGFRKTKRYETIEVYLDNRNFNKKKQVFNGFT 120

Query: 121 AQICQHELDHLEGILI 136
 AQI QHE+DH EGI+I
 50 Sbjct: 121 AQIIQHEMDHFEGLIII 136

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6019> which encodes the amino acid sequence <SEQ ID 6020>. Analysis of this protein sequence reveals the following:

55 Possible site: 45
 >>> Seems to have no N-terminal signal sequence

-2185-

INTEGRAL Likelihood = -3.61 Transmembrane 55 - 71 (55 - 73)

----- Final Results -----

5 bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 77/136 (56%), Positives = 103/136 (75%)

Query: 1 MIKPIVRDTFFLQKQSQMASRADVSLAKDLQETLHANQNYCVMGAANMIGSLKRVIIINV 60
MI+ I+ D F LQKQ+Q+A + D+ + +DLQ+TL + C+GMAANMIG KR++I+++
Sbjct: 1 MIREIITDHFLLQKQAQVAKKEDLWIGQDLQDTLAFYRQECLGMAANMIGEQRIVIVSM 60

15 Query: 61 GITNLVMFNPVNVVAKSDPYETEESCLSLVGCRSTQRYCHITISYRDINWKEQQIKLTDFP 120
G +LVMFNPV+V+K Y+T+ESCLSL G R TORY IT+ Y D NW+ +++ IT
Sbjct: 61 GFIDLVMFNPVMVSKGIYQTKESCLSLSGYRKTQRYDKITVEYLDHNWRPKRLSLTGLT 120

20 Query: 121 AQICQHLDHLEGILI 136
AQICQHLDHLEGILI
Sbjct: 121 AQICQHLDHLEGILI 136

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1942

A DNA sequence (GBSx2051) was identified in *Sagalactiae* <SEQ ID 6021> which encodes the amino acid sequence <SEQ ID 6022>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence

30

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.2880(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05820 GB:AP001514 NADP-specific glutamate dehydrogenase
[Bacillus halodurans]

40 Identities = 298/444 (67%), Positives = 362/444 (81%), Gaps = 2/444 (0%)

Query: 7 YVASVLEKVKQNEHEEEFLQAVEEVFESLVPVFDKYPOYIEENLLERLVEPERVISFRV 66
YV V E VK++N +E EF QAV+EVF+SL+PV K+PQY+++ +LER+VEPERVISFRV
Sbjct: 16 YVQHVIYETVKKRNPNEHEFHQAVKEVFDLPLVVKHPQYVVKQAILERIVEPERVISFRV 75

45 Query: 67 PWVDDKGQVQVNRGVRVQFSSAIGPYKGLRFHPTVTQSIKFLGFEQIFKNSLTGLPIG 126
PWVDD+G VQVNRG+RVQF+SA+GPYKGLRFHP+V SI+KFLGFEQIFKN+LTG PIG
Sbjct: 76 PWVDDQGNVQVNRGFRVQFNSALGPYKGLRFHPSVNASIIKFLGFEQIFKNALTGQPIG 135

50 Query: 127 GKGGSNFDPRGKSDNEVMRFTQSFMTLQKYIGPDLVDPAGDIGVGGREIGYLYGQYKR 186
GKGGS+FDPKGS D E+MRF+QSF+EL YIGPD+DVPAGDIGVG +EIGY++GQYK+
Sbjct: 136 GKGGSDFDPKGS DGEIMRFSQSFMSLSNYIGPDIDVPAGDIGVGAKEIGYMGQYKK 195

55 Query: 187 L-NGYQNGVLTCKGLTYGGSLARTEATGYGAVYFAKEMLAARGQDLTGKVALVSGSGNVA 245
+ G++ GVLTKGK+ YGGLAR EATGYG VYF +EM+ G G +VSGSGNV+
Sbjct: 196 MRGGFEAGVLTCKGIGTYGGSLARKEATGYGTVYFVEEMIKDHGFSFAGSTVVVSGSGNVS 255

Query: 246 IYATEKLQELGATVVAVSDSSGYVYDDPGIDLETLKQIKEVERARIVKYTEKHPKANFTP 305
IYA EK +LGA VVA SDS GYVD +GIDL+T+K++KEVER RI +Y +HP A++
Sbjct: 256 IYAMEKAMQLGAKVVACSDSGGYVYDKNGIDLQTVKRLKEVERKRISYVNEHPAHYVQ 315

60

Query: 306 ADQGSIIWSIKADLAFPCATQNELDEEDAKLLVENGVLAVTEGANMPSTLGAIKVFQKAGV 365

-2186-

G IWS+ D+A PCATQNELDE A +L+ NGV AV EGANMPSTL A+ FQ+ GV
 Sbjct: 316 GCSG-IWSVPCDIALPCATQNELDEAAATMLIANGVKAVGEGANMPSTLQAVHTFQEHGV 374

Query: 366 AFGPAKAAANAGGVAVSALEMAQNSSRRATFEEVDQELQRIMKTIFVNASEADEFGDSG 425
 F PAKAANAGGV+VSALEMAQNS+R AWTTFEEVD +L IMK I+ + +AA+ + SG
 Sbjct: 375 LFAPAKAANAGGVSVSALEMAQNSTRLATWTFEEVDKLYEIMKNIYRESIKAELYEASG 434

Query: 426 NLVLGANIAGFLKVAQAMSAQGIV 449
 NLV+GANIAGF+KVA AM + G+V
 Sbjct: 435 NLVVGANIAGFVKVADAMISHGVV 458

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

15 Example 1943

A DNA sequence (GBSx2052) was identified in *S.agalactiae* <SEQ ID 6023> which encodes the amino acid sequence <SEQ ID 6024>. Analysis of this protein sequence reveals the following:

Possible site: 58
 >>> Seems to have no N-terminal signal sequence

20	INTEGRAL	Likelihood = -8.55	Transmembrane	61 - 77 (55 - 87)
	INTEGRAL	Likelihood = -7.70	Transmembrane	177 - 193 (175 - 202)
	INTEGRAL	Likelihood = -7.06	Transmembrane	99 - 115 (95 - 122)
	INTEGRAL	Likelihood = -5.89	Transmembrane	42 - 58 (40 - 60)
25	INTEGRAL	Likelihood = -3.08	Transmembrane	160 - 176 (159 - 176)
	INTEGRAL	Likelihood = -2.44	Transmembrane	124 - 140 (122 - 144)

----- Final Results -----

	bacterial membrane	---	Certainty=0.4418(Affirmative)	< succ>
30	bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
	bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 9955> which encodes amino acid sequence <SEQ ID 9956> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

35 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1944

40 A DNA sequence (GBSx2053) was identified in *S.agalactiae* <SEQ ID 6025> which encodes the amino acid sequence <SEQ ID 6026>. This protein is predicted to be ABC transporter, ATP-binding protein (msbA). Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have a cleavable N-term signal seq.

45	INTEGRAL	Likelihood = -10.72	Transmembrane	152 - 168 (147 - 192)
	INTEGRAL	Likelihood = -5.47	Transmembrane	267 - 283 (264 - 288)
	INTEGRAL	Likelihood = -4.30	Transmembrane	171 - 187 (169 - 192)
	INTEGRAL	Likelihood = -2.13	Transmembrane	67 - 83 (67 - 83)
	INTEGRAL	Likelihood = -0.32	Transmembrane	493 - 509 (493 - 509)

50 ----- Final Results -----

	bacterial membrane	---	Certainty=0.5288(Affirmative)	< succ>
	bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

-2187-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:CAB69752 GB:AL137187 putative ABC transporter [Streptomyces coelicolor A3(2)]
Identities = 269/611 (44%), Positives = 392/611 (64%), Gaps = 31/611 (5%)

10 Query: 9 RLWSYLTRYKATLFLAIFLKVLSFMSILEFFILGLAITELTANLV--DMAKG----- 59
RL S +ATLF + V+S ++++ P ILG A + A +V DM G
Sbjct: 27 RLVSQFRPERATLFTLLACVVSVGLNVVGPVKILGRATDLVFAGIVGRDMPSGATKEQVL 86

Query: 60 -----VSGAELNVPYIAGILIIYFFRGVFYELGSYGSNYFMTTVV 99
V G ++ + +L++ L + + V
Sbjct: 87 ATMREHGDGNVADMLRSTDFVPGQIDFGAVGEVLLALATFAVAGLLMAVATRLVNRAV 146

15 Query: 100 QKSIRDIRHDLNRKINKVPVSFYFDKHQFGDMLGRFTSDVETVSNALQQSFLQIINAFLSI 159
+++ +R D+ K++++P+SYFDK Q G++L R T+D++ + LQQS Q+IN+ L+I
Sbjct: 147 NRTMFRLREDVQTKLSRLPLSYFDKRQGEVLSRATNDIDNIGQTLQQSMGQLINSLTI 206

20 Query: 160 ILVVVMVLYLNVPLAMIIACIPVTYFSAQAILKRSQPYFKEQAKILGELNGFVQEKLTG 219
I V+ M+ Y++ LA++ + +P+++ A + KRSQP F +Q + G+LN ++E TG
Sbjct: 207 IGVLAMMFYVSWILALVALVTVP LSFV VATR V GKRSQPQFVQQRSTGQLNAHIEEYTG 266

25 Query: 220 FNIIKLYGREEASSQEFDRITDNLRHVGFKASFISGIMFVLNSISDFIYLIIFVGGGLQ 279
++K++GR+E S+++F + D L GFKA F SGIM P++ +S+ Y+++A VGGL+
Sbjct: 267 HALVKVFGREQESAKQFAEQNDALYEAGFKAQFNSGIMQPLMMCVSNLNYVLVAVVGGLR 326

30 Query: 280 VIAGTLTIGNMQAFVQYVWQISQPVQTITQLAGVLQSAKSSLERIFEVL D-EEEEANQVT 338
V +G L+IG++QAF+QY Q S P+ + +A ++QS +S ER+FE+LD EE+ A+ +
Sbjct: 327 VASGQLSIGDVQAFIQYSRQFSMPLTQVASMALVQSGVASAERVFELLDAAEQSADPIP 386

35 Query: 339 EKLSHDLTGQVSFHGVD FHYSPDKPLIRDFNL DVEPGQMIAIVGPTGAGKTTLINLLMRF 398
DL G+V V F Y P+KPLI D +L VEPG +AIVGPTGAGKTTL+NLMLRF
Sbjct: 387 GARPEDLRGRVELEHVSFRYDPEKPLIEDLSLKVEPGHTVAIVGPTGAGKTTLVNLLMRF 446

40 Query: 399 YDVSEGAITVDGHDIRHLSRQDFRQFGMVLDQAWLYEGTIKENLRF G-NLEASDEDIVA 457
Y+VS G IT+DG DI +SR + R GMVLQD WL+ GTI EN+ +G + E + +I
Sbjct: 447 YEVSGGRITLDGVDIKMSRDEL RAGIGMVLDQTLWLFGGTIAENIAYGASREVTREIEE 506

45 Query: 458 AAKAANVDHFIRTLPGGYNMVMNQESSNISLGQKQLLTARALLADPKILILDEATSSVD 517
AA+AA+ D F+RTLP GY+ V++ E + +S G+KQL+TIARA L+DP IL+LDEATSSVD
Sbjct: 507 AARAAHADRFVRTLPDGYDTVIDDEGTGVSAGEKQLITITARAFLSDPVILVLDEATSSVD 566

Query: 518 TRLELLIQKAMKKLMGRTSFVIAHRLSTIQEADNILVLKDGQIEQGNHQKLLADKGFY 577
TR E+LIQKAM KL GRTSFVIAHRLSTI++AD ILV++DG I+EQG H +LL G Y
Sbjct: 567 TRTEVLIQKAMAKLAHGRTSFVIAHRLSTIRDADTILVMEDGAIVEQGAHTELLTADGAY 626

Query: 578 YELYSQFSNS 588
LY +QF+ +
Sbjct: 627 ARLYKAQFAEA 637

There is also homology to SEQ IDs 160 and 6546.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1945

55 A DNA sequence (GBSx2054) was identified in *S.agalactiae* <SEQ ID 6027> which encodes the amino acid sequence <SEQ ID 6028>. Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have a cleavable N-term signal seq.

60 INTEGRAL Likelihood = -10.88 Transmembrane 242 - 258 (235 - 263)
INTEGRAL Likelihood = -9.82 Transmembrane 159 - 175 (129 - 177)
INTEGRAL Likelihood = -9.71 Transmembrane 52 - 68 (49 - 77)

-2188-

INTEGRAL Likelihood = -8.49 Transmembrane 134 - 150 (129 - 158)
 INTEGRAL Likelihood = -1.17 Transmembrane 272 - 288 (272 - 289)

----- Final Results -----

5 bacterial membrane --- Certainty=0.5352(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB69751 GB:AL137187 putative ABC transporter [Streptomyces
 coelicolor A3(2)]
 Identities = 226/565 (40%), Positives = 342/565 (60%), Gaps = 1/565 (0%)

15 Query: 6 SYLKRYPNWLWLDLLGAMLFVTVILGMPTALAGMIDNGVTKGDRGTGVYLWTFIMFIFVVL 65
 ++YL+ Y + L + L L +PT A +ID GV KGD + + +M +
 Sbjct: 8 TYLRPYKKPIALILVALQFLQTCASLYLPTLNAHIIDEQVVGKDSGYILSYGALMIGISIA 67

20 Query: 66 GIIGRITMAYASSRLTTTMRDMRNDMYAKLQEQYSHHEYEQIGVSSLVTRMTSDTFVLMQ 125
 ++ I + +R + RD+R ++ ++Q +S E G SL+TR T+D +
 Sbjct: 68 QVVCNIGAVFYGARTAAALGRDVRGAVFDRVQSFSAREVGHFGAPSLITRTTNDVQQVQM 127

25 Query: 126 FAEMSLRLGLVTPMVMIFSVMILITSPSLAWLVAVAMPLLVGVLVVAIKTKPLSERQQ 185
 A M+ L + P++ + +VM L L+ ++ +P+L + + K +PL + Q
 Sbjct: 128 LALMTFTLMVSAPIMCVGGIVMALGLDVLPSGVLLGVVPVLAICVTLIVRKLRLPRKMQ 187

30 Query: 186 TMLDKINQYVRENLTGLRVVRAFARENFQSQKFQVANQRYTDTSTGLFKLTGLTEPLFVQ 245
 LD +N+ +RE +TG RV+RAF R+ ++ Q+F+ AN T+ + G L L P+ +
 Sbjct: 188 VRLDPTVNRVLRQITGNRVIRAFVRDEYEQQFRKANTELTEVALGTGNLLALMFVVM 247

35 Query: 246 IIIAMIVAIVWFALDPLQGAIKIGDLVAFIEYSFHALFSFLLFANLFTMYPRMVSSHR 305
 ++ +A+VWF + G ++IGDL AF+ Y + S ++ +F M PR V + R
 Sbjct: 248 VVNLSIAVVFGAHRIDSGMQIGDLTAFLAYLMQIVMSVMMATFMFMMPRAEVCAR 307

40 Query: 306 IREVMMPISINPNTGVTDTKLKGHLEFDNVTFAYPGETESPVLDHDSFKAKPGETIAF 365
 I+EV++ S+ P VT+ + GHLE F YPG E PVL I A+PGET A
 Sbjct: 308 IQEVLTESSVPPVAPVTELRHGHLEIREAGFRYPG-AEEPVLRHIDLVARPGETTAV 366

45 Query: 366 IGSTGSGKSSLVNLI PRFYDVTLGKILVDGVDVRDYNLKSIRQKIGFIPQKALLFTGTIG 425
 IGSTGSGKS+L+ L+PR +D T G++LV+GVDVR + K+L + + +PQK LF GT+
 Sbjct: 367 IGSTGSGKSTLLGLVPRLFDATDGEVLVNGVDVRTVDPKTLAKVSLVPQKPYLFAGTVA 426

50 Query: 426 ENLKYGKADATIDDLRQAVDISQAKFIESHQEAFETHLAEGGSNLSGGQKQRLSIARAV 485
 NL+YG DAT ++L A+ ++QAKEF+ + + +A+GG+N+SGGQ+QRL+IAR +
 Sbjct: 427 TNLRYGNPDATDEELWHALAVAQAKEFVSELEGGLDAPIAQGGTNVSGGQRLAIARTL 486

55 Query: 486 VKDPDLYIFDSSFSALDYKTATLRARLKEVTGDSTVLIVAQRVGTIMDADQIIVLDEGE 545
 V+ P++Y+FDSSFSALDY TDA LRA L + T ++TV+IVAQRV TI DAD+I+VLDEG
 Sbjct: 487 VQRPEIYLFDDSSFSALDYATDAALRAELAQETAETVIVIVAQRVATIRDADRIVVLDEGR 546

60 Query: 546 IVGRGTHAQLIENNAIYREIAESQL 570
 +VG G H +L+ +N YREI SQL
 Sbjct: 547 VVGVRHHELMADNETYREIVLSQL 571

55 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4985> which encodes the amino acid
 sequence <SEQ ID 4986>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have an uncleavable N-term signal seq

60 INTEGRAL Likelihood = -16.24 Transmembrane 155 - 171 (145 - 176)
 INTEGRAL Likelihood = -7.48 Transmembrane 130 - 146 (122 - 150)
 INTEGRAL Likelihood = -5.04 Transmembrane 13 - 29 (12 - 30)
 INTEGRAL Likelihood = -5.04 Transmembrane 56 - 72 (52 - 75)
 INTEGRAL Likelihood = -4.14 Transmembrane 239 - 255 (238 - 259)
 INTEGRAL Likelihood = -1.70 Transmembrane 269 - 285 (269 - 288)

65 ----- Final Results -----

-2189-

bacterial membrane --- Certainty=0.7496(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 175/511 (34%), Positives = 296/511 (57%), Gaps = 3/511 (0%)

Query: 59 MFIFVVLGIIGRITMAYASSRLTTTMRDMRNDMYAKLQEYSHHEYEQIGVSSLVTRMTS 118
 + I +LG++ +++++ + DMR + K+Q++S+ E +LV R+T+
 10 Sbjct: 56 LLIIALLGLMSGAINTVLAAKIAQGVSAADMREKTFRKIQDFSYANIEAFNAGNLVRLTN 115

Query: 119 DTFVLMQFAEMSLRLGLVTPMVMIFSVVMILITSPSLAWLVAVAMPLVGVILYVAITK 178
 D + M ++ P++ I + +M + T P L W++ V + L+ ++ V +
 15 Sbjct: 116 DINIQSLVMMMFQILFRLPILFIGAFIMAVQTFPQLWVIVVMVILIALINGLVMRQMG 175

Query: 179 PLSEKQQTMLDKINQYVRENLTGLRVVRAFARENFQSQKQVANQRYTDTSTGLFKLTGL 238
 P + Q ++DKIN+ +ENL G+RVV++F +E Q KF+ + + + L
 Sbjct: 176 PRFGKFQRLMDKINRIAKENLRGRVVKSFVQEQQQYTKFKETSNDLLALNLSIGYGFSL 235

Query: 239 TEPLFVQIIIIAMIVAIVWFALDPLQRGAIKIGDLVAFIEYSFHALFSFLLFANLFTMYPR 298
 +P + + + + ++ IG++ +F+ Y +FS ++ ++ R
 20 Sbjct: 236 MQPALMLVSYLAVYVSINNVSTMVETDPTVIGNIASFMTYMMQIMFSIIIVGSMGMQVSR 295

Query: 299 MVVSSHRIREVMDMPISINPNTGVTDTKLKGHLEFDNVTFAYPGETESPVLHDISFKAK 358
 VS RIR+++ ++ E + + G + FD+V+F YP + E P L ISF +
 25 Sbjct: 296 AFVSMARIRQILSTEPAMTFENE--KEETISGSIVFDDVSFTYPNDDE-PTLKHISFAIE 352

Query: 359 PGETIAFIGSTGSGKSSLVNLIPRFYDVTILGKILVDGVDVRDYNLKSIRQKIGFIPQKAL 418
 PG+ + +G+TSGSKS+L LIPR +D G+IL+ G ++ + +LRQ + + QKA+
 30 Sbjct: 353 PGQMVGIVGATGSGKSTLAQLIPRLFDPDGQILLGGKPIKTLSTTLRQSVSIVLQKAI 412

Query: 419 LFTGTIGENLKYGKADATIDDLRQAVDISQAKEFIESHQEAFETHLAEGGSNLSSGGQKOR 478
 LF+GTI +NL+ G A A ID +++A I+QAKEFI+ +E+ + E GSNLSSGGQKOR
 35 Sbjct: 413 LFSGTIADNLRQGSKADIDAMQKAAQIAQAKEFIDRMDSRYESQVEERGSNLSSGGQKOR 472

Query: 479 LSIARAVVKDPDLYIFDDSFSAIDYKTDATLRLKEVTGDSSTVLIVAQRVGTIMDADQI 538
 LSIAR V+ P + I DDS SALD K++ ++ L +T +IVAQ++ +++ AD+I
 40 Sbjct: 473 LSIARGVINHPKILILDSTSDAKSEKRVQEQALSHKLEGTITVIVAQKISSVVKADKI 532

Query: 539 IVLDEGEIVGRGTHAQLIENNAIYREIAESQ 569
 +VLD+G+++G GTHA+L+ NNAIYREI E+Q
 45 Sbjct: 533 LVLQDQQLIGEGTHAELVANNAIYREIYETQ 563

There is also homology to SEQ IDs 72 and 6552.

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1946

A DNA sequence (GBSx2055) was identified in *S.galactiae* <SEQ ID 6029> which encodes the amino acid sequence <SEQ ID 6030>. Analysis of this protein sequence reveals the following:

50 Possible site: 24
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 55 bacterial cytoplasm --- Certainty=0.2391(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA51784 GB:X73368 ORF 18.3 [Salmonella typhimurium]

-2190-

Identities = 58/162 (35%), Positives = 92/162 (55%), Gaps = 8/162 (4%)

Query: 1 MIIRPIIKNDDQAVAQLIRQSLRAYDL--DKPDTAYSDPHLDHLTSYYEKIEKSGFFVIE 58
 + +R I D+ A+A++IRQ Y L DK T +DP+LD L Y + + ++V+E
 5 Sbjet: 9 LTVRRITTADNAAIARVIRQVSAEYGLTADKGYTV-ADPNLDELQVYSQ-PGAAYWVVE 66

Query: 59 ERDEIIGCGGFGPLKNL---IAEMQKVYIAERFRGKGLATDLVKMIEVEARKIGYRQLYL 115
 + ++G GG PL I E+QK+Y RG+GLA L M AR+ G+++ YL
 10 Sbjet: 67 QNGCVVGGGGVAPLSCSEPDICELQKMYFLPVIRGQGLAKKLALMALDHAREQGFKRCYL 126

Query: 116 ETASTLSRATAVYKHMGYCALSQPIANDQGHTAMDIWMIKDL 157
 ET + L A A+Y+ +G+ +S+P+ GH ++ M+KDL
 15 Sbjet: 127 ETTAFLREAIALYERLGFHEISEPL-GCTGHVDCEVRMLKDL 167

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1947

20 A DNA sequence (GBSx2056) was identified in *S.galactiae* <SEQ ID 6031> which encodes the amino acid sequence <SEQ ID 6032>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1738(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12566 GB:Z99108 similar to ABC transporter (ATP-binding
 protein) [Bacillus subtilis]
 Identities = 269/625 (43%), Positives = 397/625 (63%), Gaps = 11/625 (1%)

35 Query: 1 MSDFLVDGLTKSVGDKTVFSNVFSFIHSLDRIGIIGVNGTGKTTLLDVISGELGFDGDRS 60
 MS + L K+ GDKT+F ++SF I +RIG+IG NGTGK+TLL VI+G +
 Sbjet: 1 MSILKAENLYKTYGDKTLFDHISFHIEENERIGLIGNGTGKSTLLKVIAGLESTIE--EG 58

40 Query: 61 PFSSANDYKIAYLKQEPDFDSDTILDTVLSSDLREMAIKEYELLNHY-----EESKQ 115
 + + ++ +L Q+P+ QT+L+ + S + M ++EYE L E +Q
 Sbjet: 59 EITKSGSVQVEFLHQDPELPAGQTVLEHIYSGESAVMKTLEYEKALYELGKDPENEQRQ 118

45 Query: 116 SRLEKVMABMSLDAWSIESEVKTVLSKLGITDLQLSVGELSGGLRRRVQLAQVLLNDAD 175
 L A+MD+ +AW + KTVLSKLG+ D+ V ELSGG ++RV +A+ L+ AD
 Sbjet: 119 KHLAQAQAKMDANNAWDANTLAKTVLSKLGVDVTKPVNELSGGQKKRVAIAKNLIQPAD 178

50 Query: 176 LLLLEPTNHLDDITIAWLTFNFKNSKKTVLFITHDRYFLDNVATRIFELDKAQITEYQG 235
 LL+LDEPTNHLDD +TI WL +L V+ +THDRYFL+ V RI+EL++ + Y+G
 Sbjet: 179 LLILDEPTNHLDDNETIEWLEGYLSQYPGAVMLVTHDRYFLNRVTNRIYELERGSLEYTYKG 238

55 Query: 236 NYQDYVRLRAEQDERDAASLHKKKQLYKQELAWMRTQPOARATKQOARINRFQNLKNDLH 295
 NY+ ++ RAE++ + K++ L ++ELAW+R +AR+TKQ+ARI+R + LK
 Sbjet: 239 NYEVFLEKRAEREAQAEQKETKRONLLRRELAWLRGAKARSTKQKARIDRVETLKEQTG 298

60 Query: 296 QTSSTSDLEMTFETSRIGKKVINFENVFSYPDKSILKDFNLLIQNKDRIGIVGDNVGVK 355
 S S L+ + R+GK+VI ENV +Y + ++ FN L+ +RIGI+G NG+GK
 Sbjet: 299 PQSSGS-LDFAIGSHRLGKQVIEAENVMIAYDGRMLVDRFNLVPIGERIGIIGFNGIGK 357

Query: 356 STLLNLIVQDLQPDSCNVSIGETIRVGYFSQQLHNMDSKRVINYLQEVADDEVKTSVGT 415
 +TLLN + PD G+++IG+T+R+GY++Q M+G +VI+Y++E A+ VKT+ G

-2191-

5 Sbjct: 358 TTLNALAGRHTPDGGDITIGQTVRIGYYTQDHSEMNGELKVIDYIKETAEVVKTAGDGM 417
 Query: 416 SVTE-LLEQFLFPRSTHGTQIAKLSGGEKKRLYLKILIEKPNVLLDEPTNDLDIATLT 474
 E +LE+FLFPRS T I KLSGGEK+RLYLL++L+++PNVL LDEPTNDLD TL+
 10 Sbjct: 418 ITAEQMLERFLFPRSMQQTIRKLSGGEKKRLYLQVLMQEPNVLFLEPTNDLDTETLS 477
 Query: 475 VLENFLQGFGGPVITVSHDRYFLDKVANKIIAFEDND-IREFFGNYYTDYLDEKAFNEQNN 533
 VLE+++ F G VITVSHDRYFLD+V ++I FE N I F G+Y+DY++E +
 15 Sbjct: 478 VLEDYIDQFPGVVITVSHDRYFLDRVDRLLIVFEGNGVISRFQGSYSYDMEESKAKKAAP 537
 Query: 534 EVISKKESTKTSREKQSRKRMSYFEKQEWATIEDDIMILENTITRIENDMQTCGSDFTRL 593
 + + +E T + K+ RK++SY ++ EW IED I LE ++E D+ GSDF ++
 Sbjct: 538 KP-AAEEKTAAEAPKKRKKLSYKDQLEWDGIEDKIAQLEEKHEQLEADIAAAGSDFGKI 596
 15 Query: 594 SDLQKELDAKNEALLEKYDRYEYLS 618
 +L E E L DR+ LS
 Sbjct: 597 QELMAEQAKTAELEAAMDRTWTELS 621

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6033> which encodes the amino acid
 sequence <SEQ ID 6034>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2591(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 467/624 (74%), Positives = 535/624 (84%), Gaps = 3/624 (0%)
 Query: 1 MSDFLVDGLTKSVGDKTVFSNVSFIIHSLDRIGIIGVNGTGKTTLLDVISGELGFGDCDRS 60
 MS FLV+ LTK+VGDKTVF ++SFIIH DRIGIIGVNGTGKTTLLDV+SG LGFDGD S
 35 Sbjct: 1 MSHFLVEKLTKTVGDKTVFQDISFIIHDFDRIGIIGVNGTGKTTLLDVLSGRLGFGDGHS 60
 Query: 61 PFSSANDYKIAYLKQEPDFDSSQTILDVTLSSDLREMALEKIYELLNHYEESKQSRLEK 120
 PFS ANDYKIAYL Q+P+F+D+ ++LDTVLS+D++ + LI++YELL+ +Y E KQ LE
 Sbjct: 61 PFSKANDYKIAYLTDPEFNDAAASVLDTVLSADVKAIQLIRQYELLMANYTEDKQESLES 120
 40 Query: 121 VMAEMDSLDAWSIESEVKTVLSKLGITDLQLSVGELSGGLRRRVQLAQVLLNDADLILLD 180
 +M+EMD LDAWSIES+VKTVLSKLGITDL+ VG+LSGG+RRRVQLAQVLL ADLILLD
 Sbjct: 121 LMSEMDRLDAWSIESDVKTVLSKLGITDLEQKVGDLSGGMRRRVQLAQVLLGAADLILLD 180
 45 Query: 181 EPTNHLIDITIAWLITNFKNSKKTVLFIITHDRYFLDNVATRIFELDKAQITEYQGNYYQDY 240
 EPTNHLIDITIAWLIT +LK +KKTVLFIITHDRYFLD+VATRIFELDKA +TEYQGNYYQDY
 Sbjct: 181 EPTNHLIDITIAWLITTYLKTAKTIVLFIITHDRYFLDHVATRIFELDKAGLTEYQGNYYQDY 240
 Query: 241 VRLRAEQDERDAASLHKKKQLYKQELAWMRTQPQARATKQARINRFQNLKNDLHQTSDT 300
 VRL+AEQDERDAA+LHKKKQLYKQELAWMRTQPQARATKQARINRF +LK ++HQ S
 50 Sbjct: 241 VRLKAEQDERDAANLHKKKQLYKQELAWMRTQPQARATKQARINRFSDLKKEVHQDSSA 300
 Query: 301 SDLEMTFETSRIGKKVINFENVSFSYPDKSILKDFNLLIQNKDRIGIVGDNGVGKSTLLN 360
 LEMTFETSRIGKKVI+FE++SF+Y D+ ++KDFNL+IQNKDRIGIVGDNGVGKSTLLN
 55 Sbjct: 301 DKLEMTFETSRIGKKVIHFEDLSFAYGDRQLIKDFNLLIQNKDRIGIVGDNGVGKSTLLN 360
 Query: 361 LIVQDLQPDSGNVSIGETIRVGYSQQLHNMDGSKRVINYLQEVADDEVKTSVGTTSVTEL 420
 +I DL+P SG + IG+TIRVGYSQQL +MD +KRVINYLQEVADDEVKTSVGTTS++EL
 Sbjct: 361 IINGDLKPTSGKLDIGDTIRVGYSQQLKMDKETKRVINYLQEVADDEVKTSVGTTSISEL 420
 60 Query: 421 LEQFLFPRSTHGTQIAKLSGGEKKRLYLKILIEKPNVLLDEPTNDLDIATLTVLENFL 480
 LEQFLFPRSHGT IAKLSGGEKKRLYLK+LIEKPNVLLDEPTNDLDIATL VLENFL
 Sbjct: 421 LEQFLFPRSSHGTLIAKLSGGEKKRLYLKLLIEKPNVLLDEPTNDLDIATLKVLENFL 480
 65 Query: 481 QGFGGPVITVSHDRYFLDKVANKIIAFEDNDIREFFGNYYTDYLDEKAFNEQNNNEVISKKE 540
 F GPVITVSHDRYFLDKVA KI+AFE+ DIR F+GNY+DYLDEK F ++ E K

-2192-

Sbjct: 481 ANFAGPVITVSHDRYFLDKVATKILAFEEGDIRVFGNYSDYLDKVFKEKTV EADLAKT 540

Query: 541 STKTS---REKQSRKMSYFEKQEWATIEDDIMILENTTTRIENDMQTCGSDFTRLSDLQ 597

+ +K+ RKRMSY EKQEW IED I +E I IEN M T SD+ +L+ LQ

Sbjct: 541 TVTEEVPLPQKEERKMSYLEKQEWAIQIEDKIATTEANIEEIEENQMLTVVSDYGQLAQLQ 600

Query: 598 KELDAKNEALLEKYDRYEYLSELD 621

KELD +N LL Y+R+EYLS LD

Sbjct: 601 KELDQRNNDLLAYERFEYLSGLD 624

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1948

A DNA sequence (GBSx2057) was identified in *S.agalactiae* <SEQ ID 6035> which encodes the amino acid sequence <SEQ ID 6036>. This protein is predicted to be poly(a) polymerase (papS). Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2658(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9957> which encodes amino acid sequence <SEQ ID 9958> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB38446 GB:L47709 poly(A) polymerase [Bacillus subtilis]

Identities = 157/395 (39%), Positives = 235/395 (58%), Gaps = 14/395 (3%)

Query: 11 FQKALPILKKIKKAGYEAYFVGGSVRDVLDRPIHDVDIATSSYPEETKQIFKRTVDVGI 70

F KALP+L+ + +AG++AYFVG+VRD + R I DVDIAT + P++ +++F+RTVDVG

Sbjct: 5 FIKALPVLRLILIEAGHQAYFVGAVRDSYMKRTIGDVDIATDAAPDQVERLFQRTVDVVGK 64

Query: 71 EHGTVLVLEKGGEYEITTFRTTEEVYVDYRRPSQVNFVRSLEEDLKRRDFTVNAFALNEDG 130

EHGT++VL + YE+TTFRTE YVD+RRPS+V F+ SLEEDLKRRD T+NA A+ DG

Sbjct: 65 EHGTIIVLWEDETYEVTTFRTESDYVDFRRPSEVQFISLEEDLKRRDLTINAMAMTADG 124

Query: 131 EVIDLFHGLDDLDNHLRAVGLASERFNEDALRIMRGLRFSASLNFDIETTTFEAMKKHA 190

+V+D F G D+D ++R VG +RF EDALR++R +RF + L F + T EA+ K

Sbjct: 125 KVLDFGGKKDIDQKVIRTVGKPEDRFQEDALRMLRAVRFMSQLGFTLSPETEATAKEK 184

Query: 191 SLEKISVERSFIEFDKLLAPYWRKGMLALIDSHAFNYLPCLKNRELQLSAFLSQLDKD 250

SLL +SVER IEF+KLL R+ + LI + + LP ++ L +S +

Sbjct: 185 SLLSHVSVERKTIEFEKLLQGRASRQALQTLIQTRLYEELPGFYHKRENL---ISTSEFP 241

Query: 251 FLFETS-EQAWASLILSMEV--EHTKTFLKKWKTSTHFQKDVEHIVDVYRIRBQMLTKE 307

F TS E+ WA+L++++ + + FLK WK K+ HI D + L

Sbjct: 242 FFSLTGREELWAALLINLIGIVLKDAPLFLKAWKLPGKVIKEAIIADTF---GQSLDAM 297

Query: 308 HLYRYGKTIKQAEQIRKAR-GLMVDFEKIEQLD---SELAIHDRHEIVVNGGTLIKLG 363

+YR GK + A I + R +D +K++ + L I ++ + G L+

Sbjct: 298 TMYRAGKALLSAAKISQLRQNEKLDEKLLKDIQYAYQNLPIKSLKDLIDITGKDLLALRN 357

Query: 364 IKPGPQMGIISQIELAIVLGQLINEEEAILHFVK 398

G + + + IE A+V G+L N+++ I ++K

Sbjct: 358 RPAGKVVSEELQWIEQAVVTGKLSNQKHHIEEWLK 392

-2193-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6037> which encodes the amino acid sequence <SEQ ID 6038>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2023(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 256/400 (64%), Positives = 312/400 (78%)

```

Query: 2  MRLNYLPSEFQKALPILKKIKAGYEAYFVGGSVRDVLDRPIHDVDIATSSYPEETKQI 61
          M+L +PSEFQKALPIL KIK+AGYEAYFVGGSVRDVL+RPIHDVDIATSSYPEETK I
Sbjct: 1  MKLMTMPSEFQKALPILTKIKEAGYEAYFVGGSVRDVLLEPPIHDVDIATSSYPEETKAI 60

Query: 62  FKRTVDVGIEHGTVLVLEKGGYEYITTFRTEEVYDYYRPSQVNFVRSLEEDLKRRDFTV 121
          F RTVDVGIEHGTVLVLE GGEYEITTFRTE++YVDYYRPSQV+FVRSLEEDLKRRDFTV
Sbjct: 61  FNRRTVDVGIEHGTVLVLENGGEYEITTFRTEDIYDYYRPSQVSFVRSLEEDLKRRDFTV 120

Query: 122  NAFALNEDGEVIDLFLHGLDDLDNHLRAVGLASERFNEDALRIMRGLRFSASLNFDIETT 181
          NA AL+E+G+VID F GL DL LRAVG A ERF EDALRIMRG RF+ASL+FDIE
Sbjct: 121  NALALDENGQVIDKFRGLIDLKQKRLRAVGKAERFEEDALRIMRGFRFAASLDFDIEAI 180

Query: 182  TFEAMKKHASLLEKISVERSFIEFDKLLAPYWRKGMLALIDSHAFNYLPCLKNRELQLS 241
          TFEAM+ H+ LLEKISVERSF EFDKLL+AP+WRKG+ A+I A++YLP LK +E L+
Sbjct: 181  TFEAMRSHSPLEKISVERSFTEFDKLLMAPHWRKGISAMIACQAYDYLPGLKQKEAGLN 240

Query: 242  AFLSQLDKDFLFETSEQAWASLILSMEVEHTKTFLLKWKSTSTHFQKDVEHIVDVYRIREQ 301
          + L +F F QAWA +++S+ +E K+FLK WKTS FQ+ V ++ +YRIR++
Sbjct: 241  HLIVSLKDNFTFSYHQAWAYVMISLAIEDPKSFLKAWKTSNDFQRYVTKLIALYRIRQE 300

Query: 302  MGLTKEHLYRYGKTIKQAEGRKARGLMVDFEKIEQLDSELAHDRHEIVVNGGTLIKK 361
          K +Y+YGK + E +RKA+ L VD ++I LD L IHD+H+IV+NG LIK
Sbjct: 301  RSFEKLDIYQYGKMASLVEDLRKAQSLSDMDRINTLDQALVIHDKHDIVLNGSHLIK 360

Query: 362  LGIKPGPQMGGDIISQIELAIVLGQLINEEAILHFKVQYL 401
          G+K GPQ+G ++ ++ELAIV G+L N+ I FV++ L
Sbjct: 361  FGMKSGPQLGLMLEKVELAIVEGRDNDFTTIEAFVREEL 400

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1949

A DNA sequence (GBSx2058) was identified in *S.galactiae* <SEQ ID 6039> which encodes the amino acid sequence <SEQ ID 6040>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2939(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB07346 GB:AP001519 unknown conserved protein [Bacillus halodurans]
Identities = 94/274 (34%), Positives = 153/274 (55%), Gaps = 2/274 (0%)

Query: 2 KLALITDTSAYLPEAIENHEDVYVLDIPIIIDGKTYIEGQNLTLTDQYYDKLAASKELPKT 61

-2194-

K+A++TD++AYL V V+ + ++ + Y E L+ +Y+KL ++LP T
 Sbjct: 3 KIAIVTDSTAYLGPKRAKELGVIVVPLSVVFGEAYQEEVELSSADFYKLEKHEEKLPPT 62
 Query: 62 SQPSLAELDDLLCQLEKEGYTHVLGLFIAAGISGFQNIQFLIEEHPNLTIAFPDTKITS 121
 SQP++ + +L KEG+ V+ + +++ ISG +Q+ + + D+ I+
 Sbjct: 63 SQPAVGLFVETFERLAKEGFEVVISIHLSSKISGTYQSALTAGSMVEGIEVIGYDSGISC 122
 Query: 122 APQGNLVRNALMCSREGMDFDVIVNKIQSQIEKIEGFIVVNDLNHLVKGGRLSNGSAIIG 181
 PQ N V A +EG D I++ + ++ VV+DL+HL +GGRL+ ++G
 Sbjct: 123 EPQANFVAEAAKLKKEGADPQTIIDHLDEVKRTNALFVVDLSHLHRGGRLNAAQLVVG 182
 Query: 182 NLLSIKPVLFHFNEEGKIVVYEKVRTEKKALKRLAEI-VKEMTADGEYDIAIHSRAQDKA 240
 +LL IKP+LHF E+G IV EKVTEKKA R+ E+ +E ++ +IH+ D A
 Sbjct: 183 SLLKIKPILHF-EDGSIVPLEKVRTEKKAWARVKELFAEEASSASSVKATVIHANRLDGA 241
 Query: 241 EQLYNLLAKAGLKDDLEIVSFGGVIATHLGEAV 274
 E+L + + D+ I FG VI THLGE++
 Sbjct: 242 EKLADERSQFSHVDVSIHFPGVIGTHLGEYSI 275

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6041> which encodes the amino acid sequence <SEQ ID 6042>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3379(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below.

Identities = 181/281 (64%), Positives = 233/281 (82%)
 Query: 1 MKLALITDTSAYLPEAIENHEDVYVLDIPIIIDGKTYIEGQNLTLQYYDKLAASKELPK 60
 MKLA+ITD++A LP ++ + ++ LDIP+IID +TY EG+NL++D +Y K+A S+ LPK
 35 Sbjct: 1 MKLAVITDSTATLPTDLKQDKAIFSLDIPVITDDETYFEGRNLSIDDFYQKMADSQNLPK 60
 Query: 61 TSQPSLAELDDLLCQLEKEGYTHVLGLFIAAGISGFQNIQFLIEEHPNLTIAFPDTKIT 120
 TSQPSL+ELD+LL L +GYTHV+GLF+A GISGFQNIQFL EEHP + +AFPDP+KIT
 40 Sbjct: 61 TSQPSLSELDNLLGLLSSKGYTHVIGLFLAGGISGFQNIQFLAEEHPETEMAFPDKIT 120
 Query: 121 SAPQGNLVRNALMCSREGMDFDVIVNKIQSQIEKIEGFIVVNDLNHLVKGGRLSNGSAII 180
 SAP G++V+N L SR+GM F I+NK+Q QI+ FI+V+DLNHLVKGGRLSNGSA++
 Sbjct: 121 SAPLGSMVKNVLDWSRQGMFQAILNKLQEQIDGTTAFIMVDDLNLHLVKGGRLSNGSALL 180
 45 Query: 181 GNLLSIKPVLFHFNEEGKIVVYEKVRTEKKALKRLAEIVKEMTADGEYDIAIHSRAQDKA 240
 GNLLSIKP+L F+EEGKIVVYEKVRTEKKA+KRL EI+ ++ ADG+Y++ IHS+AQDKA
 Sbjct: 181 GNLLSIKPILRFDEEGKIVVYEKVRTEKKAMKRLVEILNDLIADGQYNVFIHSKAQDKA 240
 Query: 241 EQLYNLLAKAGLKDDLEIVSFGGVIATHLGEAVAFGITPK 281
 + L LL +G + D+E V FG VIATHLGEGA+AFG+TP+
 50 Sbjct: 241 DYLRLLQDSGYQYDIEEVHFGAVIATHLGEGAIAFGVTPR 281

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

55 Example 1950

A DNA sequence (GBSx2059) was identified in *S.galactiae* <SEQ ID 6043> which encodes the amino acid sequence <SEQ ID 6044>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence
 60 INTEGRAL Likelihood = -1.59 Transmembrane 51 - 67 (50 - 67)

-2195-

----- Final Results -----

5 bacterial membrane --- Certainty=0.1638(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6045> which encodes the amino acid sequence <SEQ ID 6046>. Analysis of this protein sequence reveals the following:

10 Possible site: 61

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -3.19 Transmembrane 50 - 66 (49 - 67)

15 ----- Final Results -----

 bacterial membrane --- Certainty=0.2275(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20 The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 94/126 (74%), Positives = 115/126 (90%)

25 Query: 1 MEVIREQEFVNQYHYDARNLEWEEENGTPKTNFEVTFQLANRDEAAKVTSIVAVLQFVIV 60
 M+++RE+EFVNQYHYDARNLEWE+ENGTP+TNFEVTFQL ++DE K T IV+VLQFVIV
 Sbjct: 1 MLVREKEFEVNQYHYDARNLEWEKENGTPETNFEVTFQLIDKDEQQKETVIVSVLQFVIV 60

 Query: 61 RDEFVISGVISQMAHIQGRLINEPSEFSQDEVENLAAPLLEIVKRLTYEVTEIALDRPGV 120
 ++EFVISGVISQM I RL+++PSEF+Q+EVE+LAAPLL++VKRLTYEVTEIALDRPG+
 30 Sbjct: 61 KEEFVISGVISQMVRIILDRILDKPSEFTQEEVESLAAPLLDMVKRLTYEVTEIALDRPGI 120

 Query: 121 TLEFNS 126
 LEF +
 35 Sbjct: 121 HLEFKN 126

SEQ ID 6044 (GBS416) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 4; MW 17.5kDa).

GBS416-His was purified as shown in Figure 214, lane 7.

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1951

A DNA sequence (GBSx2060) was identified in *S.agalactiae* <SEQ ID 6047> which encodes the amino acid sequence <SEQ ID 6048>. Analysis of this protein sequence reveals the following:

45 Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.3875(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2196-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1952

A DNA sequence (GBSx2061) was identified in *S.agalactiae* <SEQ ID 6049> which encodes the amino acid sequence <SEQ ID 6050>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

10 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1953

A DNA sequence (GBSx2062) was identified in *S.agalactiae* <SEQ ID 6051> which encodes the amino acid sequence <SEQ ID 6052>. This protein is predicted to be PTS system, fructose-specific enzyme II, BC component (fruA-1). Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

25 INTEGRAL Likelihood = -10.56 Transmembrane 630 - 646 (618 - 653)
 INTEGRAL Likelihood = -7.43 Transmembrane 307 - 323 (303 - 331)
 INTEGRAL Likelihood = -7.01 Transmembrane 415 - 431 (412 - 435)
 INTEGRAL Likelihood = -7.01 Transmembrane 448 - 464 (444 - 474)
 INTEGRAL Likelihood = -3.72 Transmembrane 595 - 611 (591 - 612)
30 INTEGRAL Likelihood = -3.61 Transmembrane 530 - 546 (529 - 553)
 INTEGRAL Likelihood = -2.39 Transmembrane 350 - 366 (350 - 371)
 INTEGRAL Likelihood = -1.70 Transmembrane 486 - 502 (486 - 506)
 INTEGRAL Likelihood = -1.49 Transmembrane 376 - 392 (376 - 392)

----- Final Results -----

35 bacterial membrane --- Certainty=0.5225 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9959> which encodes amino acid sequence <SEQ ID 9960> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:BAB04547 GB:AP001510 PTS system, fructose-specific enzyme II, BC
 component [Bacillus halodurans]
 Identities = 320/659 (48%), Positives = 438/659 (65%), Gaps = 46/659 (6%)

 Query: 1 MKIQDLLKKEVMIMDLKATSKEAAIDEMITKLVDTGVTNFAIFKDGIMKREAQTSTGLG 60
 +KI +LLKK+ M+++L+A SKEA IDE++ L G + + FK I++RE+Q++TG+G
 Sbjct: 2 LKISELLKKDTMVLNLRAASKEAVIDELVRTLDKAGRLNDAQAFKRAILERESQSTTGVG 61

50 Query: 61 DGIAMPHSKNAAVKEATVLFKASAGVDYEALDGQPTDLFFMIAAPDGANDTHLAALAE 120
 +GIA+PH+K AAVK+ + F +S +G+DYE+LDGQP+ LFFMIAA +GAN+ HL L+ L
 Sbjct: 62 EGIAIPHAKTAAVKQPAIAFGSDAGIDYESLDGQPShLFFMIAASEGANNEHLETLSRL 121

-2197-

5 Query: 121 SKYLLKEGFADQLRQAKTPDDIIATFDSNSISQETVAPQTVQSTSKGSDYIVAVTACTTG 180
S +L+ E F L +A++ D+I+A D +E + +G + ++AVT C TG
Sbjct: 122 STFLMDETFRSTLMKAQSEDEILAAID----KKEAETAGEAEEKQEGYE-LLAVTGCPTG 176

10 Query: 181 IAHTYMAEEALKKKAEMGVGIKVTNGASGVGNKLTSSDIARAKGVIIAADKAVEMDRF 240
IAHTYMA + LK KA E+GV IKVETNG+ GV N+LT +I+ AK +I+AAD VEMDRF
Sbjct: 177 IAHTYMAADNLKSKAQELGVSIKVTNGSGGVKNRLTDEEISAAKAIIVAADTKVEMDRF 236

15 Query: 241 DGKPLVSRPVADGIKKSEDLINIILDNKAQTYHAKNQNDKQSGESDGKSGGLGS---AFYK 297
GKP++ PV DGI++ ++LI+ L KA Y + Q+ DG +G G FYK
Sbjct: 237 HGKPIVQPVTDGIRRPKELIDQALAGKAPVY----EGGAQASGEDGSAGGGRPKLGFYK 292

20 Query: 298 HLMGGVSQLPFPVIGGGIMIAIAFLFDNILGVPKDQLSNLGSYHEIAALFKNIGGA-AFA 356
HLM GVS MLPFV+GGGI+IAI+F+F P D SYH A + IGG AF
Sbjct: 293 HLMNGVSNMLPFPVVGGLILIAISFMFGIKAFDPSPD-----SYHPFAEMLMTIGGNAFG 347

25 Query: 357 FMLPVLAGYIAYSIAEKPGLVAGFVAGSIASSGLAFGKVPFAEGGKATLALAGVPSGFLG 416
M+PVL A +IA STA++PG AG + G IAS+G A GFLG
Sbjct: 348 LMIPVLAAFIAMSIAADRPFGAAGMIGGLIASTGEA-----GFLG 386

30 Query: 417 ALVGGLAGGVILLRKLKSLGLPKSLEGIKSILLYPLLGLVITGFLMLLVNIPMAAINTA 476
L+ GFLAG V L ++K+L+ LP++L+GIK+IL YP+ + ITG +ML++ P+AA NT
Sbjct: 387 GLIAGFLAGYVALGVKKVLANLQTLTGDIKILFYFVNIFITGMIMLVIVGPLAFAFNTG 446

35 Query: 477 LNTFLQGLSGSSAVLMGLLVGMMMAVDMGGPVNKAAYVFGTGLAATVANGGSVVMAAVM 536
L +L + ++ V++G+++GGMAVDMGGP+NKAA+ FG + A G AAVM
Sbjct: 447 LQDWLGSSMGTANMVLGVILGMMMAVDMGGPINKAAFTFGIAMIDA----GNFGPHAAMV 502

40 Query: 537 AGGMVPLAVFVATLLFKDKFNNEERQSGLTINVMGLSFITEGAIPFGAADPARAIPFSI 596
AGGMVPL + +AT LFK KF +ER++G TN ++G SFITEGAIPF AADP R IPS I
Sbjct: 503 AGGMVPLGIALATTLFKKKFTKQEREAGKTNILGASFITEGAIPFAAADPGRVIPSII 562

45 Query: 597 VGSALTGALVGLAGIKLMAHPGGIFVI---ALTSNPLLYILFILIGAVVSGVLFGLFRK 652
VGSA G L L + L APHGG FVI + +NPLLY++ I+ G++V+ +L G ++K
Sbjct: 563 VGSAPAGGLTALFNVTLTAPHGGAFVIFIGNIVNPLLYLVAILAGSIVTALLLGFWKK 621

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6053> which encodes the amino acid sequence <SEQ ID 6054>. Analysis of this protein sequence reveals the following:

40 Possible site: 18
>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.77	Transmembrane	624 - 640 (612 - 646)
INTEGRAL	Likelihood = -7.59	Transmembrane	301 - 317 (297 - 321)
INTEGRAL	Likelihood = -6.85	Transmembrane	442 - 458 (439 - 468)
45 INTEGRAL	Likelihood = -5.95	Transmembrane	409 - 425 (406 - 426)
INTEGRAL	Likelihood = -3.61	Transmembrane	524 - 540 (523 - 547)
INTEGRAL	Likelihood = -2.50	Transmembrane	337 - 353 (337 - 353)
INTEGRAL	Likelihood = -2.44	Transmembrane	589 - 605 (589 - 605)
INTEGRAL	Likelihood = -1.70	Transmembrane	480 - 496 (480 - 500)
50 INTEGRAL	Likelihood = -1.44	Transmembrane	370 - 386 (370 - 386)

----- Final Results -----

bacterial membrane	---	Certainty=0.5310(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
55 bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the databases:

>GP:BAB04547 GB:AP001510 PTS system, fructose-specific enzyme II, BC component [Bacillus halodurans]

60 Identities = 322/659 (48%), Positives = 431/659 (64%), Gaps = 48/659 (7%)

Query: 1 MKIQDLLRKDIMILDLOAISKEVAIDEMITKLVKDIVHDFDVFKKSIMTREEQTSTGLG 60
+KI +LL+KD M+L+L+A SKE IDE++ L + ++D FK++I+ RE Q++TG+G
Sbjct: 2 LKISELLKKDTMVLNLRASKEAVIDELVRTLDKAGRLNDAQAFKRAILERESQSTTGVG 61

65

-2198-

Query: 61 DGIAMPHSKNIVVDKPAVLFAKSNKGVVDYKALDGOPTDLFFMIAAPQGANDTHLAALAE 120
 +GIA+PH+K V +PA+ F +S+ G+DY++LDGQP+ LFFMIAA +GAN+ HL L+ L
 Sbjct: 62 EGIAPHAHTAAVKQPAIAFGRSDAGIDYESLDGQPSHLFFMIAASEGANNEHLETLSRL 121

5 Query: 121 SQYLLKDGFPADKLRAAATPEAVIAVFD--EASTAKEEVVAPTSGQDFIVAVTACPTGIAH 178
 S +L+ + F L A + + ++A D EA TA E + ++AVT CPTGIAH
 Sbjct: 122 STFLMDETFRSTLMKAQSEDEILAAIDKKEAETAGEAEEKQEGYE--LLAVTGCPTGIAH 179

10 Query: 179 TYMAEEALKQAEMGVAIKVETNGASGVANRLTAEDIQRAKGVIVAADKAVEMDRFDGK 238
 TYMA + LK +A E+GV+IKVETNG+ GV NRIT E+I AK +IVAAD VEMDRF GK
 Sbjct: 180 TYMAADNLKSKAQELGVSIVKVNRLTDEEISAAKAIIVAADTKVEMDRFHGK 239

15 Query: 239 QFIARPVADGIKKSQELISLILNNEGNTYHAKNGKSETAVSTEKTSLGG-----AFYKHL 293
 I PV DGI++ +ELI L + Y + S E S GG FYKHL
 Sbjct: 240 PVIQVPVTDGIRRPKELIDQALAGKAPVY-----EGGAQASGEDGSAGGGRPKLGFYKHL 294

20 Query: 294 MGGVSQLPFVIGGGIMIALAFLLDNMLGVPNDQLGSLGSYHEIAAIFMNIIGGA-AFSFM 352
 M GVS MLPFV+GGGI+IA++F+ P+D SYH A + M IGG AF M
 Sbjct: 295 MNGVSNMLPFVVGGLIAISFMFGIKAFDPSDF-----SYHPPAEMLMITIGGNAGFLM 349

25 Query: 353 LPVLAGYIAYSIAEKPLVAGFVAGAIASNGLAFGKVPFAAGGEVSLGLTGVPVSGFLGAL 412
 +PVL A +IA SIA++PG AG + G IAS G A GFLG L
 Sbjct: 350 IPVLAFAFIAMSIAIDRPGFAAGMIGGLIASTGEA-----GFLGGL 388

30 Query: 413 VGGFLAGGVILALRKLLAGLPRSEGVKSILLYPLLGLVLTGFLMLFVNIPMAAINTALN 472
 + GFLAG V L ++K+LA LP++L+G+K+IL YP+ + +TG +ML + P+AA NT L
 Sbjct: 389 IAGFLAGYVALGVKKVLANLPQTLTGILFYFVNIFITGMIMLVIVGPLAFAFNTGLQ 448

35 Query: 473 DFLQGLSGSSAVLMGLLVGGMMAVDMGGPVNKAAYVFGTGTAAATVANGSGSVMAAVMAG 532
 D+L + ++ V++G+++GGMMAVDMGGP+NKAA+ FG + A G AAVMAG
 Sbjct: 449 DWLGSMTANMVLIGVILGGMMAVDMGGPINKAAFTFGIAMIDA---GNFGPHAAMVAG 504

40 Query: 533 GMVPPPLAVFVATLLFKDKFTKEERESGLTNIVMGLSFITEGAIPFGAADPARAIPSFIA 592
 GMVPPPL + +AT LFK KFTK+ERE+G TN ++G SFITEGAIPF AADP R IPS I G
 Sbjct: 505 GMVPPPLGIALATTLFKKFTKQEREAGKTNILGASFTITEGAIPFAAADPGRVIPSIIVG 564

Query: 593 SALTGALVGLAGIKLMAHPGGIFVI---ALTSNEPILYLVFVIGALVSGILFGALRKKA 648
 SA G L L + L APHGG FVI + +NP+LYLV ++ G++V+ +L G +K A
 Sbjct: 565 SAFAGGLTALFNVTLSAPHGGAFFVIFIGNIVNNPLLVLVAIIAGSIVTALLLGFWKDA 623

An alignment of the GAS and GBS proteins is shown below.

Identities = 526/652 (80%), Positives = 581/652 (88%), Gaps = 6/652 (0%)

45 Query: 1 MKIQDLLKKEVMIMDLKATSKEAAIDEMITKLVDITGVVTNFAIFKDGIMKREAQTSTGLG 60
 MKIQDLL+K++MI+DL+A SKE AIDEMITKL+ +V +F +FK IM RE QTSTGLG
 Sbjct: 1 MKIQDLLRKDIMILDQAISKEVAIDEMITKLVEKDIVHDFDVFKKSIMTREEQTSTGLG 60

50 Query: 61 DGIAMPHSKNAVKEATVLFASASGVVDYALDGOPTDLFFMIAAPDGANDTHLAALAE 120
 DGIAMPHSKN V + VLFAS GVDY+ALDGOPTDLFFMIAAP GANDTHLAALAE
 Sbjct: 61 DGIAMPHSKNIVVDKPAVLFAKSNKGVVDYKALDGOPTDLFFMIAAPQGANDTHLAALAE 120

55 Query: 121 SKYLLKEGFADQLRQAKTPDDIIATFDSNSISQETVAPQTVQSTSGSDYIVAVTACTIG 180
 S+YLLK+GFAD+LR A TP+ +IA FD S ++E V T G D+IVAVTAC TG
 Sbjct: 121 SQYLLKDGFPADKLRAAATPEAVIAVFD--EASTAKEEVVAPT-----SQQDFIVAVTACTPTG 175

60 Query: 181 IAHTYMAEEALKKAAEMGVGKIKVETNGASGVGNKLTSSDIARAKGVIIAADKAVEMDRF 240
 IAHTYMAEEALKK+AAEMGV IKVETNGASGV N+LT+ DI RAKGVI+AADKAVEMDRF
 Sbjct: 176 IAHTYMAEEALKQAEMGVAIKVETNGASGVANRLTAEDIQRAKGVIVAADKAVEMDRF 235

65 Query: 241 DGKPLVSRPVADGIKKSSEDLINIILDNKAQTYHAKNQNDKQSGESDGKSLGSAFYKHL 300
 DGK ++RPVADGIKKS++LI++IL+N+ TYHAKN ++ S K+ LG AFYKHL
 Sbjct: 236 DGKQFIARPVADGIKKSQELISLILNNEGNTYHAKN-GKSETAVSTEKTSLGGAFYKHL 294

Query: 301 GGVSQLPFVIGGGIMIAIAFLFDNILGVPKDQLSNLGSYHEIAALFKNIGGAFAFMLP 360
 GGVSQLPFVIGGGIMIA+AFI DN+LGVP DQL +LGSYHEIAA+F NIGGAFAF+FMFLP
 Sbjct: 295 GGVSQLPFVIGGGIMIALAFLLDNMLGVPNDQLGSLGSYHEIAAIFMNIIGGAFAF+FMFLP 354

-2199-

Query: 361 VLAGYIAYSIAEKPLVAGFVAGSIASSGLAFGKVPFAEGGKATLALAGVPSGFLGALVG 420
 VLAGYIAYSIAEKPLVAGFVAG+IAS+GLAFGKVPFA GG+ +L L GVPSPGFLGALVG
 Sbjct: 355 VLAGYIAYSIAEKPLVAGFVAGAIASNGLAFGKVPFAAGGEVSLGLTGVPSPGFLGALVG 414

5 Query: 421 GFLAGGVILLRKLKLLSGLPKSLEGIKSILLYPLLGVLTITGFLMLLVNIPMAAINTALNTF 480
 GFLAGGVIL LRKLL+GLP+SLEG+KSILLYPLLGVLT+TGFLML VNIPMAAINTALN F
 Sbjct: 415 GFLAGGVILALRKLKLLAGLPRSLGKVSILLYPLLGVLTGFLMLFVNIPMAAINTALNDF 474

10 Query: 481 LQGLSGSSAVLMGLLVGGMMMAVDMGGFVNKAAYVFGTGTLAATVANGGSSVMAAVMAGGM 540
 LQGLSGSSAVLMGLLVGGMMMAVDMGGFVNKAAYVFGTGTLAATVANGGSSVMAAVMAGGM
 Sbjct: 475 LQGLSGSSAVLMGLLVGGMMMAVDMGGFVNKAAYVFGTGTLAATVANGGSSVMAAVMAGGM 534

15 Query: 541 VPPLAVFVATLLFKDKFNNEERQSGLTNIVMGLSFITEGAIPFGAADPARAIPSFIVGSA 600
 VPPLAVFVATLLFKDKF EER+SGLTNIVMGLSFITEGAIPFGAADPARAIPSF I GSA
 Sbjct: 535 VPPLAVFVATLLFKDKFTKEERESGLTNIVMGLSFITEGAIPFGAADPARAIPSF IAGSA 594

20 Query: 601 LTGALVGLAGIKLMAHPGGIFVIALTSNPLLYILFILIGAVVSGVLFGLFRK 652
 LTGALVGLAGIKLMAHPGGIFVIALTSNP+LY++F++IGA+VSG+LFG RK
 Sbjct: 595 LTGALVGLAGIKLMAHPGGIFVIALTSNPILYLVFVVGALVSGILFGALRK 646

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1954

A DNA sequence (GBSx2063) was identified in *S. agalactiae* <SEQ ID 6055> which encodes the amino acid sequence <SEQ ID 6056>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.1532(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:AAC24914 GB:AF012285 fructose-1-phosphate kinase [Bacillus subtilis]
 Identities = 146/303 (48%), Positives = 197/303 (64%)

Query: 1 MIYTVTLNPSIDFIVRLDTLLGSGVNRMTSDDKYVGGKGINVSRILKRLKIDNTATGFIG 60
 MIYTVTLNPS+D+IV ++ +G +NR + D KY GKGGINVSR+LKR + + A GF+G
 40 Sbjct: 1 MIYTVTLNPSVDYIVHVEDFTVGGLNRSSYDTKYPGGKGINVSRLLKRHHVASKALGFVG 60

Query: 61 GFTGHFVEDGLVLEGIKTDFVSVNEDTRINVVKAKIETEINGGGPRITNEQLHRLEKLL 120
 GFTG +++ L E ++T F V DTRINVK+K ETEING GP I++E +
 45 Sbjct: 61 GFTGEYIKTFLREENLETAPEVKGDTTRINVKLKTGDETEINGQGPTISDEDFKAFLEQF 120

Query: 121 SRLTPEDTVVFAGSAPASLGKNKYNTLIPIAKKTGAEVVCFEGQTLDDALAYQPLLVKP 180
 L D VV AGS P+SL + Y + K+ A VV D G+ LL A +P L+KP
 50 Sbjct: 121 QSLQEGDIVVLGSIPLPHDTYEKIAEACKQONARVVLDISGEALLKATEMKPFLMKP 180

Query: 181 NNHELADIFGVELEGLPDIEKYAHKILDKGAKNVIVSMAGDGALLVTPEASYFAKPIKGE 240
 N+HEL ++FG + + + Y K+++GA++VIVSMAGDGALL T EA YFA KG+
 55 Sbjct: 181 NNHELGEFMGTATTSVEEAVPYGKKLVEQGAHVIVSMAGDGALLFTNEAVFANVPKKG 240

Query: 241 VNNSVGAGDSMVAGFTGEFVKSNNPVEALKWGVACGTATTFSDDLATAEFIQDIYNKVEV 300
 + NSVGAGDS+VAGF K EA + GV G+AT FS++L T EF+Q + +V+V
 60 Sbjct: 241 LVNSVGAGDSVVAGFLAGISKQLPLEEAFRLGVTSGSATAFSEELGTEEFVQQLLPEVKV 300

Query: 301 EKL 303
 +L
 60 Sbjct: 301 TRL 303

-2200-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6057> which encodes the amino acid sequence <SEQ ID 6058>. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1738(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 222/302 (73%), Positives = 261/302 (85%)

```

Query: 1  MIYTVTLNPSIDFIVRLDTLLGSVNRMTSDDKYVGGKGINVSRILKRLKIDNTATGFIG 60
          MIYTVTLNPSIDFIVR+D + LGSVNRM SDDK+ GGKGINVSRIL+RL I +TATGF+G
Sbjct: 1  MIYTVTLNPSIDFIVRIDQINLGSVNRMASDDKFAGGKGINVSRILQRLDIATATGFLG 60

Query: 61  GFTGHFVEDGLVLEGIKTDFVSVNEDTRINVKVKAKIETEINGGGPRITNEQLHRLEKLL 120
          GFTG F+E+ L EG+KTDFV ++DTRINVK+K++ ETE+NG GP I+ EQL L+ L
Sbjct: 61  GFTGRFIEESLSAEGVKTDVFKGDQDTRINVKIKSQEETELNGQGPISQEQLEDLKTCL 120

Query: 121  SRLTPEDTVVFAGSAPASLGKQVNTLIPIAKKTGAEVVCDFFEGQTLDDALAYQPLLVKP 180
          S+LT EDTVVFAGSAPA+LGN VY L+P+ +++GA+VVCDFFEGQTL+DALAY PLLVKP
Sbjct: 121  SQLTAEDTVVFAGSAPANLGNNAVYKELLPLVRQSGAQVCDFFEGQTLIDALAYNPLLVKP 180

Query: 181  NNHELADIFGVELEGLPDIEKYAHKILDKGAKNVIVSMAGDGALLVTPEASYFAKPIKGE 240
          NNHEL IFG L L D+E YA ++L+ GA+NVI+SMAGDGALLVT EA+YFAKPIKGE
Sbjct: 181  NNHELEAIFGTILTSLDDVETYARRLLEMGAQNVIISMAGDGALLVTKEATYFAKPIKGE 240

Query: 241  VKNSVGAGDSMVAGFTGEFVKSKNPVEALKWGVACGTATTFSDDLATAEFIQDIYNKVEV 300
          VKNSVGAGDSMVAGFTGEF+KS+NP+EALKWGVACGTAT FSDDLAT FI++ Y+KVEV
Sbjct: 241  VKNSVGAGDSMVAGFTGEFMKSONPIEALKWGVACGTATATFSDDLATIAFIKETYHKVEV 300

Query: 301  EK 302
          EK
Sbjct: 301  EK 302

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1955

A DNA sequence (GBSx2064) was identified in *S.agalactiae* <SEQ ID 6059> which encodes the amino acid sequence <SEQ ID 6060>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2769(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9961> which encodes amino acid sequence <SEQ ID 9962> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC24913 GB:AF012285 FruR [Bacillus subtilis]

Identities = 97/247 (39%), Positives = 148/247 (59%), Gaps = 4/247 (1%)

Query: 23 MLKSKRKEIILSRLEQNKSVTLDELTSILETSESTVRRDLDELESAGFLKRVHGGALPY 82

-2201-

ML +R ++I+ ++E++ V + EL ++ SEST+RRDL LE GFLKRVHGA
 Sbjct: 1 MLTPERHQLIIDQIEKHVVVKIQELINLTNASESTIRRDLSLEERGFLKRVHGGAAKLS 60

Query: 83 SLGQELSNQEKAIKNVQKKLDIARQTAKLIAKQDVIFIDAGTTTELLIDFLPH-EQLTVV 141
 + E EK+ KN+ KL IA + A L+ + D I++DAGITT +IDF+ + + VV
 Sbjct: 61 DIRLEPDMLEKSSKNLHDKLKAIEKAASLLEEGDCIYLDAGTTTLHMDIFMDKTKDIVVV 120

Query: 142 TNSIHHAALKLVDRGIKTIIGGAVKHSTDASIGQVAINQIRQITVDKAFGLMNGID-EVY 200
 TN + H L+ + I ++GG VKH T A IG ++ + Q DK+FLG NG+ E
 Sbjct: 121 TNGVMHIDALIRKEISFYLLGGYVKHRTGAIIGGASLVAMDQYRFDKSFLGTNGVHTEAG 180

Query: 201 LITPDLEEAIAKEAIINNSQQTFFILMDSSKIGQVTFKVKKEINDINLVNKTDSSELMTII 260
 TTPD +EA +K+ I ++ ++L D SK G+++F+ I D ++T TD+E +T
 Sbjct: 181 FTTPDPDEALLKQKAIKQAKHAYVLADPSKFGEISFSFAFAGIGDATIIT--TDAEELTFD 238

Query: 261 KEKMKVI 267
 + K +
 Sbjct: 239 NYQEKTV 245

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6061> which encodes the amino acid sequence <SEQ ID 6062>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2604(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below.

Identities = 135/237 (56%), Positives = 184/237 (76%)

Query: 33 LSRLEQNKSVTLDELTSILETSESTVRRDLDELESAGFLKRVHGGAEPLPYSLGQELSNQE 92
 ++++ + V+L++L +L +SEST+RRDL ELE G L RVHGGAE L +SL +ELSNQE
 35 Sbjct: 1 MAKITEENYVSLDLMQLNNSSESTIRRDLEGELEQEGRLHRVHGGAE L FHS LQEELSNQE 60

Query: 93 KAIKNVQKKLDIARQTAKLIAKQDVIFIDAGTTTELLIDFLPHEQLTVVNSIHHAALKV 152
 K++KN K IA++ ++LI DVIFIDAGTTTE L+ FL + LTVVNSIHHAALV
 40 Sbjct: 61 KSVKNSHIKKAIQASQLIYDNDVIFIDAGTTTEFLLPFLQAKNLTVVNSIHHAARLV 120

Query: 153 DRGIKTIIGGAVKHSTDASIGQVAINQIRQITVDKAFGLMNGIDEVYLTTPDLEEAIAK 212
 + I+TII+GG VK +TDASIG VA+ QIRQ+ DKAFLGMNG+D+ YLTTPD+EEA IK
 45 Sbjct: 121 ELSIETIIVGGYVKQTTDASIGNVALEQIRQMNFDKAFGLMNGVDDSYLTTPDMEEAVIK 180

Query: 213 EAIINNSQQTFFILMDSSKIGQVTFKVKKEINDINLVNKTDSSELMTIIKEKMKVIQV 269
 +A+++N++ +IL+D +KIGQV+F KV IND+ ++T + ++ IKEK KVI++
 50 Sbjct: 181 KAVLSNAKLAYILVDGTFKIGQVSFVKVAPINDVTITLGGASILKQIKEKAKVIEL 237

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1956

A DNA sequence (GBSx2065) was identified in *S.agalactiae* <SEQ ID 6063> which encodes the amino acid sequence <SEQ ID 6064>. This protein is predicted to be beta-lactam resistance factor. Analysis of this protein sequence reveals the following:

55 Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5777(Affirmative) < succ>

-2202-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:CAB89121 GB:AJ277485 beta-lactam resistance factor
 [Streptococcus pneumoniae]
 Identities = 215/410 (52%), Positives = 283/410 (68%)

10 Query: 1 MTLRELTIEEFKEHSGNYDSQSFLQTPEMAKLLEKRGYDVRVLYGYQVENKLEIISLSYIM 60
 M L LT EEF+ +S S+SF+Q+ +M LLEKRG + YL + E ++++ +L Y +
 Sbjct: 1 MALTTLTKEEFQTYSDQVSSRSFMQSVQMGDLLEKRGARIVYLALKQEGEIQVAALVYSL 60

15 Query: 61 PVTGGFQMKIDSGPVHSNSKYLKQFYKALQGYAKSNGVLELIVEPYDDYQLFTSSGVPSN 120
 P+ GG M+++SGP+++ L FY L+ YAK NGVLEL+V+PY+ YQ F S G P +
 Sbjct: 61 PMLGGLHMEINSGPIYTTQDQDALPVMFYAELKEYAKQNGVLELLVKPYETTYQTFDSQGNPID 120

20 Query: 121 QGNDNLIEDFTSSGYHHDGLTGTFTGKYLSTWHYVKNLEGVTSSETLLSSFSKTRALVKKA 180
 ++I+D T GY DGLTTG+ G W Y K+L +T ++LL SFSK G+ LVKKA
 Sbjct: 121 AEKKSIIQDLTDLGYQFDGLTTGYPGGEEDWLYYKDLTELTEKSLKSFSSKKGKPLVKKA 180

25 Query: 181 MSFGIKVRVLKRDELHLFKEITTSTSNRRDYMDKSLDYYQDFYDSFEGKAEFVIATLNFR 240
 +FGI+++ LKR+EL +FK IT TS RR+Y DKSL+YY+ FYD+F +AEF+IA+LNF
 Sbjct: 181 ETPGIRLKKLKEELSIFKNITKETSERREYSKSLLEYEHFYDTFGEQAEFLIASLNFS 240

30 Query: 241 EYDHNLQIKAEALENKLKLLDERFRFNADSPKYHRQRSEIINQLASFETRREQEVQSFIQK 300
 +Y LQ + LE L L +N S K Q E +Q +FE R+ E + I+K
 Sbjct: 241 DYMSKLQGEQSKLEENLDKLRDLKSNPHSEKKQNQLREYSSQFETFEVRKAEARDLIEK 300

35 Query: 301 YDNQDVVLAGSLFVYSLKETVYFFSGSYTEFNKFYAPAVLQEVYVMEALKRGSTFYNNLG 360
 Y +D+VLAGSLFVY +ET Y FSGSYTEFNKFYAPA+LQ+YVM E++KRG YN LG
 Sbjct: 301 YGEEDIVLAGSLFVYMPQETTTLFSGSYTEFNKFYAPALLQKYVMLESIKRGIPKYNFLG 360

Query: 361 IQGTFDGSIDSILRFKQNFNGCIIRKMGTFFNYPSPFKYKGIQLLKKVLKR 410
 IQG FDGSD +LRFKQNFNG I+RK GTF Y+PSP KYK IQLLKK++ R
 Sbjct: 361 IQGIFDGSIDGVLRFKQNFNGYIVRKAGTFRYHPSPLKYKAIQLLKKIVGR 410

There is also homology to SEQ ID 5460.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 1957

A DNA sequence (GBSx2066) was identified in *S.agalactiae* <SEQ ID 6065> which encodes the amino acid sequence <SEQ ID 6066>. This protein is predicted to be cell wall protein, 40 kDa (sr 5' region). Analysis of this protein sequence reveals the following:

45 Possible site: 42
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.45 Transmembrane 25 - 41 (23 - 42)

----- Final Results -----

50 bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9963> which encodes amino acid sequence <SEQ ID 9964> was also identified.

55 The protein has homology with the following sequences in the GENPEPT database.

!GB:AF278686 choline binding protein D; CbpD [Strept...
 !GB:AF278686 choline binding protein D; CbpD [Strept...

-2203-

>GP:AAF87768 GB:AF278686 choline binding protein D; CbpD
[Streptococcus pneumoniae]
Identities = 63/230 (27%), Positives = 108/230 (46%), Gaps = 34/230 (14%)

5 Query: 324 WTEQGGQDDIKWYTAVTTGDG-----NYKVAVSFADHKNEKGLYNIHLVYQEASGTLVG 377
W+ G + W + V GD NY S+ + +++++ G VG
Sbjct: 123 WSTAGTYGHVAVVSNVM-GDQIEIEEYNYGYTESYNKRVIKANTMTGFIHFKDLDDGGSVG 181

10 Query: 378 VTGKTQVTVAGTNSSQEPIENGLAKTGVYNIIGSTEVEKNEAKISSQTQFTLEKGDKINYDQ 437
+ + + GT+ + + +K E S G+K++YDQ
Sbjct: 182 NSQSTSTGGTHYFKT-----KSAIKTEPLASGTVIDYYPGEKVHYDQ 225

15 Query: 438 VLTADGYQWISYKSYSGVRRYIPVKKLTTSSSEKAKDEATKPTSYPNLPKTG-TYFTTKTV 496
+L DGY+W+SY +Y+G RY+ ++ + + P L TG T+ F
Sbjct: 226 ILEKDGKWLSTAYNGSYRYVQLEAVNKN-----PLGNSVLSSTGGTHYFKTKS 275

20 Query: 497 DVKSQPKVSSPVEFNFQKGEKIHYDQVLVDGHQWISYKSYSGIRRYIEI 546
+K++P VS+ V + GEK+HYDQ+L DG++W+SY +Y+G RRYI++
Sbjct: 276 AIKTEPLVSATVIDYYPGEKVHYDQILEKDGKWLSTAYNGSRRYIQL 325
Identities = 49/161 (30%), Positives = 85/161 (52%), Gaps = 14/161 (8%)

25 Query: 116 GNYVYSKETEVKNTPSKSAPVAFYAKKGDVFDQVFNKDNVWISYKSFVCGVRRYAAIE 175
G + + ++ +K P S V Y G+KV YDQ+ KD KW+SY ++ G RY +E
Sbjct: 191 GTHYFKTKSAIKTEPLASGTVIDYYPGEKVHYDQILEKDGKWLSTAYNGSYRYVQLE 250

30 Query: 176 SLDPGGGSETKAPTPTVNSGNNQEKIATQGNITFSHKVEVKNEAKVASPTQFTLDKGR 235
+++ + P+ NS + +T G + F K +K E V++ G++
Sbjct: 251 AVNKN-----PLGNSVLS-----STGGTHYFKTKSAIKTEPLVSATVIDYYPGEK 296

35 Query: 236 IFYDQILITIEGNQWLSYKSFNGVRRFVLLGKASSVEKTEDK 276
+ YDQIL +G +WLSY ++NG RR++ L +S + +++
Sbjct: 297 VHYDQILEKDGKWLSTAYNGSRRYIQLLEGVTSSQNYQNQ 337
Identities = 52/192 (27%), Positives = 90/192 (46%), Gaps = 13/192 (6%)

40 Query: 295 ISNETTTGFDILITINIKDDNGIAAVKVPVWTEQGGQDDIKWYTAVTTGDGNYKVAVSFAD 354
I T TGF + KD +G + T GG K +A+ T ++
Sbjct: 161 IKANIMTGF----IHFKDLDDGGSVGNSSSTSTGGTHYFKTKSAIKTEPLASGTVIDYY- 215

45 Query: 355 HKNEKGLYNIHLY---YQEASGTLVGVGTGKTQVTVAGTNSSQEPIENGLAKT--GVYNIIG 409
+ EK Y+ L Y+ S T + V+ N+ P+ N+ + G+
Sbjct: 216 YPGEKVHYDQILEKDGKWLSTAYNGSYRYVQLEAVNKN--PLGNSVLSSTGGTHYFKT 273

50 Query: 410 STEVEKNEAKISSQTQFTLEKGDKINYDQVLTADGYQWISYKSYSGVRRYIPVKKLTTSS 469
+ +K E +S+ G+K++YDQ+L DGY+W+SY +Y+G RRYI ++ + TSS+
Sbjct: 274 KSAIKTEPLVSATVIDYYPGEKVHYDQILEKDGKWLSTAYNGSRRYIQLLEGV-TSSQ 332

55 Query: 470 KAKDEATKPTSY 481
++++ +SY
Sbjct: 333 NYQNQSGNISSY 344
Identities = 33/113 (29%), Positives = 56/113 (49%), Gaps = 2/113 (1%)

60 Query: 91 NTATKDITITPLVETKPMVEKTLPEQQNYVYSK-ETEVKNTPSKSAPVAFYAKKGDVFDY 149
N + + + V P+ L G Y K ++ +K P SA V Y G+KV YD
Sbjct: 241 NGSYRYVQLEAVNKNPLGNSVLSSTGGTHYFKTKSAIKTEPLVSATVIDYYPGEKVHYD 300

65 Query: 150 QVFNKDNVWISYKSFVCGVRRYAAIESLDPGGGSETKAPTPTVNSGNNQEKI 202
Q+ KD KW+SY ++ G RRY +E + S + ++ +++ GS++ +
Sbjct: 301 QILEKDGKWLSTAYNGSRRYIQLLEGVTSSQNYQNQSGN-ISSYGHSSSTV 352

A related GBS gene <SEQ ID 8937> and protein <SEQ ID 8938> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
McG: Discrim Score: -6.74
GvH: Signal Score (-7.5): 1.26
Possible site: 42

-2205-

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the databases:

>GP:AAF87768 GB:AF278686 choline binding protein D; CbpD
 [Streptococcus pneumoniae]

10

Identities = 93/217 (42%), Positives = 136/217 (61%), Gaps = 18/217 (8%)

Query: 42 GDNYPKWKKGNG-IDSWNMYIRQCTSF AAFRLSSANGFQLPKGYGNACTWGHIKNQGY 100
 GD+YP+ +K G+ ID W MY RQCTSF AFRLS+ NGF++P YGNA WGH A+ +GY
 Sbjct: 51 GDDYPAYYKNGSQEIDQWRMYSRQCTSFVAFRLSNVNGFEIPAAAYGNANEWGHRARREGY 110

15

Query: 101 PVNKTGPSIGAIWFDKNAYQSNAAIGHVAVWADIRGDTVTIEEYNNAGQGPERYHKRQI 160
 V+ TP+IG+I W + YGHVAVW+++ GD + IEEYNY E Y+KR I
 Sbjct: 111 RVDNTPITIGSITW-----STAGTYGHVAVWSNVMGDQIEIEEYNYGY---TESYNKRVI 161

20

Query: 161 PKSQVSGYIHFKDLSSQTSYPRQLKHISQASFDPSGTYHFTTRLPVKGQTSIDSPDLA 220
 + ++G+IHFKDL + + SQ+S GT++F T+ +K + +
 Sbjct: 162 KANTMTGFIHFKDLDGSGVGN-----SQSSTSTGGTHYFKTKSAIKTEPLASGTVID 213

25

Query: 221 YYEAGQSVYYDKVVTAGGYTWLSYLSFSGNRRYIPIK 257
 YY G+ V+YD+++ GY WLSY +++G+ RY+ ++
 Sbjct: 214 YYYPGKVVHYDQILEKDGKWLSTAYNGSYRYVQLE 250

An alignment of the GAS and GBS proteins is shown below.

Identities = 34/94 (36%), Positives = 52/94 (55%)

30

Query: 453 SGVRRYIPVKKLTTSSEKAKDEATKPTSYPNLPKTGTFTTKTVDVKSQPKVSSPVEFNF 512
 S V YI K L++ + + K S + +GTY FT + VK Q + SP +
 Sbjct: 163 SQVSGYIHFKDLSSQTSYPRQLKHISQASFDPSGTYHFTTRLPVKGQTSIDSPDLAYY 222

35

Query: 513 QKGEKIHYDQVLVVDGCHQWISYKSYSGIRRYIEI 546
 + G+ ++YD+V+ G+ W+SY S+SG RRYI I
 Sbjct: 223 EAGQSVYYDKVVTAGGYTWLSYLSFSGNRRYIPI 256
 Identities = 30/78 (38%), Positives = 45/78 (57%), Gaps = 2/78 (2%)

40

Query: 402 TGVYNIIGSTEIVKNEAKISSQTQFTLEKGDKINYDQVLTADGYQWISYKSYSGVRRYIPV 461
 +G Y+ VK + I S E G + YD+V+TA GY W+SY S+SG RRYIP+
 Sbjct: 197 SGTYHFTTRLPVKGQTSIDSPDLAYYEAGQSVYYDKVVTAGGYTWLSYLSFSGNRRYIPI 256

45

Query: 462 KKLITSSEKAKDEATKPT 479
 K+ + +++ TKP+
 Sbjct: 257 KE--PAQSVVQNDNTKPS 272
 Identities = 27/94 (28%), Positives = 47/94 (49%)

50

Query: 198 NQEKIATQGNITFSHKVEVKNEAKVASPTQFTLDKGDRIFYDQILTEGNQWLSYKSFNG 257
 +Q G Y F+ ++ VK + + SP + G ++YD+++T G WLSY SF+G
 Sbjct: 190 SQASFDPSGTYHFTTRLPVKGQTSIDSPDLAYYEAGQSVYYDKVVTAGGYTWLSYLSFSG 249

55

Query: 258 VRRFVLLGKASSVEKTEDEKVPQPQARITKTG 291
 RR++ + + + D K S + +T G
 Sbjct: 250 NRRYIPIKEPAQSVVQNDNTKPSIKVGDITVTFPG 283
 Identities = 23/73 (31%), Positives = 35/73 (47%)

60

Query: 103 ETKPMVEKTLPEQGNVYSKETEVKNTPSKSAPVAFYAKKGDKVFYDQVFNKDNVWISY 162
 + K + + + G Y ++ VK S +P Y + G V+YD+V W+SY
 Sbjct: 185 QLKHSQASFDPSGTYHFTTRLPVKGQTSIDSPDLAYYEAGQSVYYDKVVTAGGYTWLSY 244

Query: 163 KSFCGVRRYAAIE 175
 SF G RRY I+
 Sbjct: 245 LSFSGNRRYIPIK 257

-2206-

SEQ ID 8938 (GBS91) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 18 (lane 7; MW 63kDa).

The GBS91-His fusion product was purified (Figure 195, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 283), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1958

A DNA sequence (GBSx2067) was identified in *S.agalactiae* <SEQ ID 6069> which encodes the amino acid sequence <SEQ ID 6070>. This protein is predicted to be thiamine biosynthesis protein. Analysis of this protein sequence reveals the following:

Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0984(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB49673 GB:AJ248285 PROBABLE 2-DEHYDROPANTOATE 2-REDUCTASE (EC
1.1.1.169) [Pyrococcus abyssi]
Identities = 85/301 (28%), Positives = 150/301 (49%), Gaps = 7/301 (2%)

Query: 1 MLVYIAGSGAMGCRFGYQISKTNHVDILLDNWADHIMAIKENGLKVTGDTEDLVKLPIMK 60
M +YI G+GA+G FG ++ DV+L+ H+ AI E GLK+ G + VK+
Sbjct: 1 MKIYILGAGAIGSLFGGLLANAGEDVLLIGR-DPHVSAINELKGLKIVGIKDLNVKVEATT 59

Query: 61 PTDATEEADLIILFTKAMQLPNMLQDIKKIIGKTKVLCLLNGLGHEDVIRQYIPEHNIL 120
E+ DLI+L TK+ L+ + I+ K + VL + NG+G+ED I ++ +
Sbjct: 60 RVPE-EKPDILIVLATKSYSTIEALKSARHIV-KGSWVLSIQNGIGNEDKIEF--GGKAI 115

Query: 121 MGVTVWTAGLKGPGCHAHLEGVGSVNLQSIDPNNQEHGHRVTELLNEAKLQATYDENVLPN 180
G+T A ++ PG G G + ++ +V ++ N A ++ EN++
Sbjct: 116 GGIITNGAMVEAPGVIKWTGKGVTTIIGLYPQGEKFKIEKVADVFNSADIETHVSENIISW 175

Query: 181 IWRKACVNGTMNSTCALLDCTIGQLFASEDGVNMVHEIIEHFVTVGKAEGVDELDEEITK 240
IW KA VN +N LL+ + ++ ++M E++ E V G+E D +
Sbjct: 176 IWAKAIVNSAINPIGTLLLEVKNKVIRENDFLLSMAMEVVKEGCRVALQNGIEFDVPPMDL 235

Query: 241 YVMDTSVKAAHHYPSMHQDLVQNQLTEIDFLNGAVNKKGENLGIDTPYCRITQLIHTKE 301
+ T + +Y SM QD+ + ++ TE+D++NG + + + + ++ P L+ LI KE
Sbjct: 236 F-FQLEQTREYNVSMQLQDIWRGKK-TEVDYINGKIVEYAKAVNLEAPMNLILLWGLIKKE 294

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6071> which encodes the amino acid sequence <SEQ ID 6072>. Analysis of this protein sequence reveals the following:

Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1392(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-2207-

Identities = 262/307 (85%), Positives = 288/307 (93%)

Query: 1 MLVYIAGSGAMGCRFGYQISKTNHDVILLDNWADHIMAIKENGLKVTGDTEDLVKLPIMK 60
 MLVYIAGSGAMGCRFGYQISKTN+DVILLDNW DHI AIKENGL VTGD E+ VKLPIMK
 5 Sbjct: 1 MLVYIAGSGAMGCRFGYQISKTNNDVILLDNWEDHINAIKENGLVVTGDVEETVKLPIMK 60

Query: 61 PTDATEEADLIILFTKAMQLPNMLQDIKKIIGKKTIVLCLLNLGHEDVIRQYIPEHNIL 120
 PT+AT+EADLIILFTKAMQLP MLQDIK IIGK+TKVLCLLNLGHEDVIRQYIPEHNIL
 10 Sbjct: 61 PTEATQEADLIILFTKAMQLPQMLQDIKGIIGKETKVLCLLNLGHEDVIRQYIPEHNIL 120

Query: 121 MGVTVWVTAGLKGPGHAHLEGVGSVNLQSIDPNNQEAGHRVTELLNEAKLQATYDENVLPN 180
 MGVTVWVTAGL+GPG AHL+GVG++NLQS+DP+NQEAGH+V +LLNEA L ATYDENV+PN
 Sbjct: 121 MGVTVWVTAGLEGPGRHQLGVGALNLQSMGPSNQEAGHQVADLLNEANLNATYDENVVVPN 180

Query: 181 IWRKACVNGTMNSTCALLDCTIGQLFASEDGVNMVHEIIEHFVTVGKAEGVELDEEITK 240
 IWRKACVNGTMNSTCALLDCTIG+LFASEDG+ MV EIIHEFV VG+AEGVEL+BBEIT+
 15 Sbjct: 181 IWRKACVNGTMNSTCALLDCTIGELFASEDGLKMKVKEIIEHFVIVGQAEGVELNBBEITQ 240

Query: 241 YVMDTSVKAHHYPSMHQDLVQNRLTEIDFLNGAVNKKGENLGIDTPYCRITQLIHTK 300
 YVMDTSVKAHHYPSMHQDLVQN RLTEIDF+NGAVN KGE LGI+TPYCR+IT+L+H K
 20 Sbjct: 241 YVMDTSVKAHHYPSMHQDLVQNRLTEIDFLNGAVNTKGBKLGINTPYCRMITELVHAK 300

Query: 301 ENVLISIK 307
 E VL+I+
 25 Sbjct: 301 EAVLNIQ 307

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1959

30 A DNA sequence (GBSx2068) was identified in *S.agalactiae* <SEQ ID 6073> which encodes the amino acid sequence <SEQ ID 6074>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 35 INTEGRAL Likelihood = -3.03 Transmembrane 61 - 77 (61 - 78)
 INTEGRAL Likelihood = -1.33 Transmembrane 80 - 96 (79 - 96)

----- Final Results -----
 bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1960

A DNA sequence (GBSx2069) was identified in *S.agalactiae* <SEQ ID 6075> which encodes the amino acid sequence <SEQ ID 6076>. This protein is predicted to be regulatory protein (pfoS/R). Analysis of this protein sequence reveals the following:

50 Possible site: 49
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.82 Transmembrane 317 - 333 (304 - 335)
 INTEGRAL Likelihood = -7.64 Transmembrane 187 - 203 (183 - 217)
 INTEGRAL Likelihood = -5.26 Transmembrane 24 - 40 (18 - 44)
 55 INTEGRAL Likelihood = -5.04 Transmembrane 143 - 159 (139 - 161)

-2208-

INTEGRAL	Likelihood = -2.34	Transmembrane	116 - 132 (115 - 136)
INTEGRAL	Likelihood = -2.13	Transmembrane	55 - 71 (55 - 71)
INTEGRAL	Likelihood = -0.96	Transmembrane	268 - 284 (268 - 284)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema
 pallidum]
 Identities = 138/358 (38%), Positives = 220/358 (60%), Gaps = 18/358 (5%)

15 Query: 2 TNTVTPKETAGSFINKVLGGTATAIVVALIPNAILATFLKPFLSYG-LAAEFLHIVQVFQ 60
 T +++P++ F+ K+L G++ IV+ L+P AI + L A H+V Q
 Sbjct: 3 TQSLSPRQ----FMMKILNGSSAGIVIGLVPPAIGELFRALAPLSPLFAALYHVVLPIQ 58

20 Query: 61 FFTPIMAGFLIGQQFKFTPMQQLAVGGAAYIGSGAWAYTEVIQKGVATGSFQLRGIGDLI 120
 F P + G L+G QF + + + + I SG + G++ + GIGD+I
 Sbjct: 59 FSVPALIGTLVGLQFHCSAPEVATLAFVSVIASG-----NVTLQNGAWLITGIGDVI 110

25 Query: 121 NMMLTAALAVLAVKWFNGKFGSLTIILLPIIIGTGVGYLGWKLPLPYVSVYVTTLIGQGINS 180
 N+ML +ALA++ V+ K GSLTII LP+I+ G +G LPYV +T +G+ I +
 Sbjct: 111 NVMLISALAIILVRALRGKLGSLTIIALPVTIVAVVAGGVGSFSLPYVKMITLFVGRVIAT 170

30 Query: 181 FTTLQPIAMSIILAMAFSMLIVSPISTVAIGLAIGLNGMSASAASMGVASTTAVLVWATM 240
 F LQP+ MSIL++M+FS++I+SP+S+VA+G+A+GL G+++ AA++GV+S L+ TM
 Sbjct: 171 FIALQPLILMSILLSMSFSLIIISPVSSVAVGIAVGLTGLASGAANIGVSSCAMTLIVGTM 230

35 Query: 241 KANKSGVPPIAIALGAMKMMMPNFKHPVMAIPMLMTATVSSLTVPFLFKLVGTPASSGFG 300
 + NK GVP+A+ GAMKM+MPN++++P++ IP+L+ V + LF L GTPAS+GFG
 Sbjct: 231 RVNKIGVPLAMFAGAMKMLMPNWIROPILNIPLLNLGLVCGVLAWLNLQGTASAGFGF 290

40 Query: 301 VGAVGPIASFEE--AGASML---IVILSWLVIPFAVGFVSHKICKDILKLYKDDIFVFE 353
 +G VGPI ++ A M+ I+ L + V+ F ++ I D LKLY+ ++F+ E
 Sbjct: 291 IGLVGPINAYRLMAYTPMVRAGILFLVYFVLSFLAAYLIDFILVDRLKLYRRELFPIE 348

There is also homology to SEQ ID 1280.

40 A related GBS gene <SEQ ID 8939> and protein <SEQ ID 8940> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8
 McG: Discrim Score: -7.24
 GvH: Signal Score (-7.5): -2.94
 Possible site: 49

45 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 7 value: -9.82 threshold: 0.0

INTEGRAL	Likelihood = -9.82	Transmembrane	317 - 333 (304 - 335)
INTEGRAL	Likelihood = -7.64	Transmembrane	187 - 203 (183 - 217)
INTEGRAL	Likelihood = -6.37	Transmembrane	143 - 159 (136 - 161)
INTEGRAL	Likelihood = -5.26	Transmembrane	24 - 40 (18 - 44)
INTEGRAL	Likelihood = -2.34	Transmembrane	116 - 132 (115 - 136)
INTEGRAL	Likelihood = -2.13	Transmembrane	55 - 71 (55 - 71)
INTEGRAL	Likelihood = -0.96	Transmembrane	268 - 284 (268 - 284)
PERIPHERAL	Likelihood = 0.69		205

55 modified ALOM score: 2.46

*** Reasoning Step: 3

60 ----- Final Results -----
 bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

10

15

20

25

30

35

40

There is also homology to SEQ ID 1276

45

Example 1961

A DNA sequence (GBSx2070) was identified in *S.agalactiae* <SEQ ID 6077> which encodes the amino acid sequence <SEO ID 6078>. Analysis of this protein sequence reveals the following:

50

55

The protein has homology with the following sequences in the GENPEPT database.

60

-2210-

5
 Query: 5 IYDITIVGGGPFVGLFAAFYAGLRGVSVKIIESLSELGGQPAILYPEKKIYDIPGYPVITG 64
 +YDITI+GGGP GLFAAFY G+R VKIIES+ +LGGQ A LYPEK IYD+ G+P +
 Sbjct: 7 LYDITIIGGGPTGLFAAFYGGMRQAKVKIIESMPQLGGQLAALYPEKYIYDVAGFPKVK 66

10
 Query: 65 RELIDKHIEQLERFKDSIEICLKEEVLSFEK-VDDVFTIQTQDKDQHLRAIVFACNGAF 123
 ++L++ Q E+F +I L++ V + K DD FTI+TDK+ H S+AI+ G GAF
 Sbjct: 67 QDLVNDLKRQAEQFNPTI--ALEQSVQNVTKETDDTFTIKTDKETHYSKAIITAGAGAF 124

15
 Query: 184 RRDAFRAHEHSVDILKASGVRILTPYVPIGLNGDSQRVSSLVQKVGDEVIELPLDNL 243
 RRD FRAHEHSV++L+ S V ILTP+ L+GD +++ + +Q+VKGD V L +D +I
 Sbjct: 185 RRDKFAHEHSVELLQKSSVNILTPFAISELSGDGEKIHVHTIQEVKGDVAVETLDVDEVI 244

20
 Query: 244 VSFSGFSTSNKNLRYWNLDYKRSSINVSSLFETTQEGVYAIGDAANYPGKVELIATGYGEA 303
 V+FGF +S ++ W L+ +++SI V++ ET G+YA GD YPGKV+LIATG+GEA
 Sbjct: 245 VNFQFVSSLGPIKQWGLEIEKNSIVVNTKMETNIPGIYAAGDICTYPGKVLIATGFG 304

25
 Query: 304 PVAINQAINIYIPDRDNRVHSTSL 328
 P A+N A +I P HSTSL
 Sbjct: 305 PTAVNNAKAFIDPTARVFGHSTSL 329

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6079> which encodes the amino acid sequence <SEQ ID 6080>. Analysis of this protein sequence reveals the following:

30
 Possible site: 20
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -0.37 Transmembrane 8 - 24 (8 - 24)

35
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

40
 >GP:CAB15201 GB:Z99120 similar to thioredoxin reductase [Bacillus subtilis]
 Identities = 173/328 (52%), Positives = 223/328 (67%), Gaps = 4/328 (1%)

45
 Query: 4 KAYDITIIGGGPIGLFAAFYAGLRGVTVKIIESLSELGGQPAILYPEKMIYDIPAYPSLT 63
 K YDITIIGGGP+GLF AFY G+R +VKIIESL +LGGQ + LYPEK IYD+ +P +
 Sbjct: 6 KVIDITIIGGGPVGLFTAFYGGMRQASVKIIESLPQLGGQLSALYPEKYIYDVAGFPKIR 65

50
 Query: 64 GVELTENLIKQLSRFEDRTTICLKEEVLTFFDKVKGK-FSIRTNKAHFSAKIIACNGA 122
 EL NL +Q+++F+ TICL++ V + +K G F + K K I NGNA
 Sbjct: 66 AQELINNLKEQMAKFDQ--TICLEQAVESVEKQADGVFKLVQMKPPTLKRSCITAGNGA 123

55
 Query: 123 FAPRTLGLSEENFADHNLFFNVHQLDQFAGQKVVICGGGDSAVDWALALEEDIAESVTVV 182
 F PR L LE+ E + NL Y V L +FAG++V I GGGDSAVDWAL LE IA+ V+++
 Sbjct: 124 FKPRKLELENAEQYEGKNLHYFVDDLQKFAGRRVAILGGGDSAVDWALMLEPIAKEVSII 183

60
 Query: 183 HRRDAFRAHEHSVELLKASTVNLLTPYVPKALKGIGNLAEKLVQKVEDEVLELELDSL 242
 HRRD FRAHEHSVE L AS VN+LTP+VP L G + E+LV+++VK D LE+D L
 Sbjct: 184 HRRDKFRAHEHSVENLHASKVNVLTFFVPAELIGEDKI-EQLVLEEVKGRKETEIDDL 242

65
 Query: 243 IVSFGFSTSNKNLKNWNLDYKRSSITVSPLFQTSQEGIFAIGDAAAYNGKVDLIATGFGE 302
 IV++GF +S +KNW LD +++SI V +T+ EG FA GD Y GKV+LIA+GFGE
 Sbjct: 243 IVNYGFVSSLGPIKNWGLDIEKNSIVVKSTMETNIEGFFAAGDICTYEGKVNLIASGFGE 302

Query: 303 APTAVNQAINIYIPDRDNRVHSTSLID 330
 APTAVN A Y+ P + +HSTSL +
 Sbjct: 303 APTAVNNAKAYMDPKARVQPLHSTSLFE 330

65 An alignment of the GAS and GBS proteins is shown below.

-2211-

Identities = 242/324 (74%), Positives = 279/324 (85%)

Query: 6 YDITIVGGGPFVGLFAAFYAGLRGVSVKIIESLSELGGQPAILYPEKKIYDIPGYPVITGR 65
 YDITI+GGGP+GLFAAFYAGLRGV+VKIIESLSELGGQPAILYPEK IYDIP YP +TG
 5 Sbjct: 6 YDITIIGGGFIGLFAAFYAGLRGVTVKIIESLSELGGQPAILYPEKMIYDIPAYPSLTGV 65

Query: 66 ELIDKHIEQLERFKDSTIEICLKKEEVLSEFEKVDVFTIQTDKQHLRSRAIVFACGNGAFAP 125
 EL + I+QL RF+D ICLKEEVL+F+KV F+I+T+K +H S+AI+ ACGNGAFAP
 10 Sbjct: 66 ELTENLIKQLSRFEDRTTICLKKEEVLTFDKVKGFSIRTNKAHFPSKATIIACGNGAFAP 125

Query: 126 RLLGLENEENYADNNLFYNVTKLEQFAGKHVVICGGGDSAVDWANELD KIAASVAIVHRR 185
 R LGLE+EEN+AD+NLFYNV +L+QFAG+ VVICGGGDSAVDWA L+ IA SV +VHRR
 Sbjct: 126 RTLGLESEENFADHNLFYNVHQLDQFAGQKVVICGGGDSAVDWALAEEDIAESVTVVHRR 185

Query: 186 DAFRAHEHSVDILKASGVRLTPYVP IGLNGDSQRVSSLVVQKVKGDEVIELPLDNLIVS 245
 DAFRAHEHSV++LKAS V +LTPYVP L G LV+QKVK DEV+EL LD+LIVS
 15 Sbjct: 186 DAFRAHEHSVELLKASTVNLITPYVPKALKGIGNLAEKLVIQKVEDEVLELELDLIVS 245

Query: 246 FGFSTSNKNLRYWNL DYKRSSINVSSLFETTQEGVYAIGDAANYPGKVELIATGYGEAPV 305
 FGFSTSNKNL+ WNL DYKRSSI VS LF+T+QEG++AIGDAA Y GKV+LIATG+GEAP
 20 Sbjct: 246 FGFSTSNKNLKNWNL DYKRSSITVSPLFQTSQEGIFAIGDAAAYNGKVDLIATGFGGEAPT 305

Query: 306 AINQAINIYIPDRDNRVHSTSLI 329
 A+NQAINIYIPDRDNRVHSTSLI
 25 Sbjct: 306 AVNQAINIYIPDRDNRVHSTSLI 329

SEQ ID 6078 (GBS178) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 5; MW 37.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 8; MW 62.4kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1962

A DNA sequence (GBSx2071) was identified in *S.agalactiae* <SEQ ID 6081> which encodes the amino acid sequence <SEQ ID 6082>. This protein is predicted to be tRNA methyltransferase (trmD). Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 40 bacterial cytoplasm --- Certainty=0.1496(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:BAB06198 GB:AP001515 tRNA methyltransferase [Bacillus halodurans]
 Identities = 144/246 (58%), Positives = 186/246 (75%), Gaps = 6/246 (2%)

Query: 2 MKIDILTLFPPEMFAPLEHS-IVGKAKERGLLEINYHNFRENAE-KSRHVDDEPYGGGQGM 59
 MKID LTLFPPEMF + HS I+ +A+ERG + NFRE +E K + VDD PYGGG GM
 50 Sbjct: 1 MKIDFLTLFPPEMFQVLHSSILKQAQERGAVSFRVNVFREYSENKHKKVDVDPYGGGAGM 60

Query: 60 LLRAQPIFDITIDKIDAQKA---RVILLDPAGRTFDQDFAEELKEDELIFICGHYEGYDE 116
 +L QP+FD ++ + + + RVIL+ P G TF Q AEEL++ + LI +CGHYEGYDE
 Sbjct: 61 VLSPPQLFDAVEDLTKKSSSTPRVILMCPQGETFTQRKAELQAELHLLCGHYEGYDE 120

55 Query: 117 RIKS-LVTDEVSLGDFVLTGGELAAMTMVDATVRLIPEVIGKETSHQDDSFSSGLLEYPQ 175
 RI+S LVTDE+S+GD+VLTGGEL AM + D+ RL+P V+G ETS Q DSFS+GLLEYPQ
 Sbjct: 121 RIRSYLVTDLSIGDYVLTGGELGAMVIADSVTRLLPAVLGNETSAQTDSFSTGLLEYPQ 180

-2212-

Query: 176 YTRPYDYLGMTVPDVLMSGHHENIRKWRLEQSLRKTLERRPDLLLENYAMTDEERLILEKI 235
 YTRP D+ G VPDVL+SGHH+NI +WR EQSL++TLERRPDLLE +T+EE+ +L+ I
 Sbjct: 181 YTRPADFRGWKVPDVLMSGHHQNIERWRKEQSLRKTLERRPDLLLEGRKLTETEEQELLSI 240

5 Query: 236 KTEIER 241
 + + E+
 Sbjct: 241 RKQQEK 246

10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6083> which encodes the amino acid sequence <SEQ ID 6084>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2705(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

20 Identities = 195/240 (81%), Positives = 224/240 (93%)

Query: 2 MKIDILTLFFEMFAPLEHSIVGKAKERGLLEINYHNFRENAEKSRHVDDEPYGGGQGMML 61
 MKIDILTLFFEMFAPLEHSIVGKAKE+GLL+I+YHNFR+ AEK+RHVDDEPYGGGQGMML
 25 Sbjct: 1 MKIDILTLFFEMFAPLEHSIVGKAKEKGLLDIHYHNFREDYAEKARHVDDEPYGGGQGMML 60

Query: 62 RAQPIFDITIDKIDAQKARVILLDPAGRTFDQFAEELSKEDELIFICGHYEGYDERIKSL 121
 RAQPIFDTI++I+A+K R+ILLDPAG+ F Q +AEEL+ E+ELIFICGHYEGYDERIK+L
 30 Sbjct: 61 RAQPIFDTIEQIEAKPRIILLDPAGKPFTQAYAEELALEEELIFICGHYEGYDERIKTL 120

Query: 122 VTDEVSLGDFVLTGGELAAMTMVDATVRLIPEVIGKETSHQDDSFSSGLLEYPQYTRPYD 181
 VTDE+SLGDFVLTGGELAAMTMVDATVRLIP+V+GKE+SHQDDSFSSGLLEYPQYTRPYD
 35 Sbjct: 121 VTDEISLGDFVLTGGELAAMTMVDATVRLIPQVLGKESHSQDDSFSSGLLEYPQYTRPYD 180

Query: 182 YLGMTVPDVLMSGHHENIRKWRLEQSLRKTLERRPDLLLENYAMTDEERLILEKIKTEIER 241
 Y GMTVPDVLMSGHHE IR WRLE+SL+KT RRPDLLE+Y ++EER +L+KIK +++
 Sbjct: 181 YRGMTVPDVLMSGHHERIRLWRLEESLKKTYLRRPDLLHYNFSEERKLLDKIKEALDQ 240

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 1963

A DNA sequence (GBSx2072) was identified in *S.agalactiae* <SEQ ID 6085> which encodes the amino acid sequence <SEQ ID 6086>. This protein is predicted to be 16S rRNA processing protein. Analysis of this protein sequence reveals the following:

45 Possible site: 43
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.71 Transmembrane 32 - 48 (32 - 52)

----- Final Results -----
 50 bacterial membrane --- Certainty=0.2084(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9381> which encodes amino acid sequence <SEQ ID 9382> was also identified.

55 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13475 GB:Z99112 similar to hypothetical proteins [Bacillus subtilis]

-2213-

Identities = 88/174 (50%), Positives = 128/174 (72%), Gaps = 1/174 (0%)

Query: 54 VTMEYFNVGKIVNTQGLQGEMRVLSTDFVEERFKKGQVLALFDEKNQFVMDIEIASHRK 113
 +T +FNVGKIVNT G++GE+RV+S TDF EER+K G L LF + +++ + +HR
 5 Sbjet: 1 MTKRWFNVGKIVNTHGIKGEVRVISKTDFAEERYKPGNTLYLFMDGRNEPVEVTVNTHRL 60

Query: 114 QKNFDIIKFKGMYHINDIEKYKGFTLKVAEDQLSDLKDGEFYHHEIIGLDVYEGE-ELIG 172
 K F +++FK ++N++E+ K +KV E++L +L +GEFY+HEIIG +V+ E ELIG
 10 Sbjet: 61 HKQFHLLQFKERQNLNEVEELKNAIKVPEBELGELNEGEFYFHEIIGCEVFTEEGELIG 120

Query: 173 KIKEILQPGANDVWVVERHGKRDLLLPYIPPVVLEVDLSNQRVQVELMEGLDDE 226
 K+KEIL PGANDVWV+ R GK+D L+PYI VV +D+ +++++ELMEGL DE
 Sbjet: 121 KVEILTPGANDVWVIGRKGDALIPYIESVVKHIDVREKKIEIELMEGLIDE 174

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6087> which encodes the amino acid sequence <SEQ ID 6088>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2787(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 133/172 (77%), Positives = 153/172 (88%)

Query: 56 MEYFNVGKIVNTQGLQGEMRVLSTDFVEERFKKGQVLALFDEKNQFVMDIEIASHRKQK 115
 MEYFNVGKIVNTQGLQGEMRVLSDFAEERFKKG LALFD+K++FV ++ I SHRKQK
 30 Sbjet: 1 MEYFNVGKIVNTQGLQGEMRVLSTDFVEERFKKGSQLALFDDKDRFVQEVTIVSHRKQK 60

Query: 116 NFDIIKFKGMYHINDIEKYKGFTLKVAEDQLSDLKDGEFYHHEIIGLDVYEGEELIGKIK 175
 +FDIIKFK MYHIN IEKYKG+TLKV++D DL++GEFYH+IIG+ VYE + LIG +K
 35 Sbjet: 61 HFDIIKFKDMYHINAIEKYKGYTLKVSKDNQDLQGEFYHQTIGMAVYEKDVLIHV 120

Query: 176 EILQPGANDVWVVERHGKRDLLLPYIPPVVLEVDLSNQRVQVELMEGLDDED 227
 EILQPGANDVW+V+R GKRDLLLPYIPPVVL VD+ N+RV VELMEGLDDED
 Sbjet: 121 EILQPGANDVWIVKRQGRDLLLPYIPPVVLNVDVFNKRVDELMEGLDDED 172

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1964

A DNA sequence (GBSx2073) was identified in *S.agalactiae* <SEQ ID 6089> which encodes the amino acid sequence <SEQ ID 6090>. This protein is predicted to be similar to *E. coli* ykfc (11). Analysis of this
 45 protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3488(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9909> which encodes amino acid sequence <SEQ ID 9910>
 55 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-2214-

>GP:AAC38715 GB:AF030367 maturase-related protein [Streptococcus pneumoniae]
Identities = 366/425 (86%), Positives = 396/425 (93%)

5 Query: 12 MSELLDKILSRNNMLEAYKQVKSNGKSAGINGVTIEQMDYLDHQNWRBTKQLIKERSYKP 71
MS+LLDKILSR NMLEAY QVKSNGKSAGI+G+TIE+MD+YL QNWR TK+LIK+R YKP
Sbjct: 1 MSKLLDKILSRNNMLEAYNQVKSNGKSAGIDGMTIEEMDNYLRQNWRLTKELIKQKRYKP 60

10 Query: 72 QPVLKVEIPKPNNGGVRNLGIPTAMDRMIQQAIVQVLSPLCEKHFSEYSYGFRPNRSCETA 131
QPVL+VEIPKP+GG+R LGIPT MDRMIQQAIVQV+SP+CE HFS+ SYGFRPNRSCE A
Sbjct: 61 QPVLKVEIPKPDGGIRQLGIPTVMDRMIQQAIVQVMSPICEPHFSYDGFRPNRSCEKA 120

15 Query: 132 IVQLLEYLNDGYEWIVDIDLEKFFDTVPQDRLMSLVHNIQDGDTESLIRKYLHSGVVIN 191
I++LLEYLNDGYEWIVDIDLEKFFDTVPQDRLMSLVHNI+DGDTESLIRKYLHSGV+IN
Sbjct: 121 IMKLLLEYLNDGYEWIVDIDLEKFFDTVPQDRLMSLVHNIIEDGDTESLIRKYLHSGVIIN 180

20 Query: 192 GQRHKTIVGTPQGGNLSPLLSNIMLNELDKGLEKRLRFVRYADDCVITVGSEAAAKRVM 251
GQR+KTIVGTPQGGNLSPLLSNIMLNELDK LEKRLRFVRYADDCVITVGSEAAAKRVM
Sbjct: 181 GQRYKTLVGTTPQGGNLSPLLSNIMLNELDKLEKRLRFVRYADDCVITVGSEAAAKRVM 240

25 Query: 252 HSVSSYIEKRLGLKVNMTKTKIIVRPNKLYLGFGFWKSPKGWKCRPHQDSVQSFKRKLKQ 311
+SVS +IEKRLGLKVNMTKTKI RP +LKYLGFQFWKS GWK RPHQDSV+ FK KLK+
Sbjct: 241 YSVSRFIEKRLGLKVNMTKTKITRPRELKYLGFGFWKSSDGWKS RPHQDSVRRFKLKLK 300

30 Query: 312 LTMKWSIDLITRIERLNWVIRGWINYFSLGNMKSIMTQIDERLRTRIRVIWQWKKKA 371
LT RKWSIDL RIE+LN IRGWINYFSLGNMKS+ IDERLRTR+R+IIWQWKKK+
Sbjct: 301 LTQRKWSIDLTRIEQLNLSIRGWINYFSLGNMKSIVASIDERLRTRLRMIWQWKKKS 360

35 Query: 372 KRLWGLLLKLGVARWIADKVSQWGDHYQLVAQKSVLKRAISKPALAKRGLVSCLDYYLERH 431
+RLWGLLLKLG +WIADKVSQWGDHYQLVAQKSVLKRAISK P L KRLVSCLDYYLERH
Sbjct: 361 RRLWGLLLKLGVPKWIADKVSQWGDHYQLVAQKSVLKRAISKPVLEKRLVSCLDYYLERH 420

Query: 432 ALKVS 436
ALKVS
Sbjct: 421 ALKVS 425

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1965

40 A DNA sequence (GBSx2074) was identified in *S.agalactiae* <SEQ ID 6091> which encodes the amino acid sequence <SEQ ID 6092>. Analysis of this protein sequence reveals the following:

Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -0.37 Transmembrane 7 - 23 (7 - 23)
45 ----- Final Results -----
bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
50

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 821> which encodes the amino acid sequence <SEQ ID 822>. Analysis of this protein sequence reveals the following:

Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
55 INTEGRAL Likelihood = -2.87 Transmembrane 1157 -1173 (1157 -1174)
----- Final Results -----
bacterial membrane --- Certainty=0.2147(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
60 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 1031/1064 (96%), Positives = 1042/1064 (97%)

5	Query: 1	MRKKQKL PFDKLAIALISTSILLNAQSDIKANTVTEDTPATEQAVEPPQPIAVSEESPSS 60
	Sbjct: 1	LRKKQKL PFDKLAIALMSTSIILLNAQSDIKANTVTEDTPATEQAVETPQPTAVSEEAPSS 60
10	Query: 61	KETKTSQTSPSDVGETVADDANDLAPQAPAKTADTPATSKATIRDLNDPSHVKTLEKAGK 120
	Sbjct: 61	KETKT QTP D ET+ADDANDLAPQAPAKTADTPATSKATIRDLNDPS VKTLEKAGK 120
15	Query: 121	GVGTVVAVIDAGFDKNHEAWRLTDKTKARYQSKENLEKAKKEHGITYGEWVNDKVAYYHD 180
	Sbjct: 121	GAGTVVAVIDAGFDKNHEAWRLTDKTKARYQSKEDLEKAKKEHGITYGEWVNDKVAYYHD 180
20	Query: 181	YSKDGKNAVDQEHGTHVSGILSGNAPSEMKEPYRLEGAMPEAQLLLMRVEIVNGLADYAR 240
	Sbjct: 181	YSKDGK AVDQEHGTHVSGILSGNAPSE KEPYRLEGAMPEAQLLLMRVEIVNGLADYAR 240
25	Query: 241	NYAQAIRDAVNLGAKVINMSFGNAALAYANLPDETKKAFDYAKSGVSIIVTSAGNDSSFG 300
	Sbjct: 241	NYAQAI DAVNLGAKVINMSFGNAALAYANLPDETKKAFDYAKSGVSIIVTSAGNDSSFG 300
30	Query: 301	GKPRPLADHPDYG VVGTPAAADSTLTVASYS PDKQLTETATVKTDDHQDKEMPVLSTNR 360
	Sbjct: 301	GKRLPLADHPDYG VVGTPAAADSTLTVASYS PDKQLTETATVKT D QDKEMPVLSTNR 360
35	Query: 361	FEPNKAYDYAYANRGTKEDDFKDVGEKIALIERGDIDFKDKIANAKKAGAVGVLIYDNQD 420
	Sbjct: 361	FEPNKAYDYAYANRG KEDDFKDV+GKIALIERGDIDFKDKIANAKKAGAVGVLIYDNQD 420
40	Query: 421	KGFPIELPNVDQMPAAFI SRDGLLLKDNPKQKITTFNATPKVLPTASGTKLSRFSSWGLT 480
	Sbjct: 421	KGFPIELPNVDQMPAAFI SRDGLLLKENPKQKITTFNATPKVLPTASGTKLSRFSSWGLT 480
45	Query: 481	ADGNIKPDIAAPGQDILSSVANNKYAKLSGTSMSAPLVAGIMGLLQKQYETQYPMTPSE 540
	Sbjct: 481	ADGNIKPDIAAPGQDILSSVANNKYAKLSGTSMSAPLVAGIMGLLQKQYETQYPMTPSE 540
50	Query: 541	RLDLAKKVLMSATALYDEDEKAYFSRQQGAGAVDAKKASAATMYVTDKNTSSKVHLN 600
	Sbjct: 541	RLDLAKKVLMSATALYDEDEKAYFSRQQGAGAVDAKKASAATMYVTDKNTSSKVHLN 600
55	Query: 601	NVSDKFEVTVTVHNKSDKPQELYYQTVQTDKVDGKHFALAPKALYETSWQKITIPANSS 660
	Sbjct: 601	NVSDKFEVTVTVHNKSDKPQELYYQATVQTDKVDGKLFALAPKALYETSWQKITIPANSS 660
60	Query: 661	KQVTVPIDASRFSKDLLAQMKNGYFLEGFVRFKQDPTKEELMSIPYIGFRGDFGNLSALE 720
	Sbjct: 661	KQVT+PID S+FSKDLLA MKNGYFLEGFVRFKQDPTKEELMSIPYIGFRGDFGNLSALE 720
65	Query: 721	KPIYDSKDGSSYYHEANSDAKDQLDGDGLQFYALKNNFTALTESNPWTIIKAVKEGVEN 780
	Sbjct: 721	KPIYDSKDGSSYYHEANSDAKDQLDGDGLQFYALKNNFTALTESNPWTIIKAVKEGVEN 780
	Query: 781	IEDIESSEITETIFAGTFAKQDDDSHYIYHRHANGKPYAAISPNGDGNRDYVQFGTFLR 840
	Sbjct: 781	IEDIESSEITETIFAGTFAKQDDDSHYIYHRHANGKPYAAISPNGDGNRDYVQFGTFLR 840
	Query: 841	NAKNLVAEVL DKEGNVVTSEVTEQVVKNYNNDLASTLGSTRFEKTRWDGKNKDGKVVAN 900
	Sbjct: 841	NAKNLVAEVL DKEGNVVTSEVTEQVVKNYNNDLASTLGSTRFEKTRWDGK+KDGKVVAN 900
	Query: 901	GTYTYRVRVYTPISSGAKEQHTDFDVIDNTTPEVATSATFSTEDSRLTLASKPKTSQPVY 960
	Sbjct: 901	GTYTYRVRVYTPISSGAKEQHTDFDVIDNTTPEVATSATFSTEDRRLTLASKPKTSQPVY 960

-2216-

Query: 961 RERIAYTYMDEDLPTEYISPNEGDTFTLPEEAETMEGATVPLKMSDFTYVVEDMAGNIT 1020
 RERIAYTYMDEDLPTEYISPNEGDTFTLPEEAETMEGATVPLKMSDFTYVVEDMAGNIT
 Sbjet: 961 RERIAYTYMDEDLPTEYISPNEGDTFTLPEEAETMEGATVPLKMSDFTYVVEDMAGNIT 1020

Query: 1021 YTPVTKLLEGHHSNKPEQDGSQAPDKKPEAKPEQDGSQGTPDKK 1064
 YTPVTKLLEGHHSNKPEQDGSQAPDKKPE KPEQDGSQG PDKK
 Sbjet: 1021 YTPVTKLLEGHHSNKPEQDGSQAPDKKPEAKPEQDGSQGAPDKK 1064

- 10 A related GBS gene <SEQ ID 8941> and protein <SEQ ID 8942> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 10
 McG: Discrim Score: 5.69
 GvH: Signal Score (-7.5): -3.33
 Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 1 value: -0.37 threshold: 0.0
 INTEGRAL Likelihood = -0.37 Transmembrane 7 - 23 (7 - 23)
 PERIPHERAL Likelihood = 2.81 508
 modified ALOM score: 0.57

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- SEQ ID 8942 (GBS276) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 2; MW 123kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 5; MW 46.5kDa).

The GBS276-His fusion product was purified (Figure 206, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 296), which confirmed that the protein is immunoaccessible on GBS bacteria.

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1966

A DNA sequence (GBSx2075) was identified in *S.galactiae* <SEQ ID 6093> which encodes the amino acid sequence <SEQ ID 6094>. Analysis of this protein sequence reveals the following:

Possible site: 58
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4286(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2217-

Example 1967

A DNA sequence (GBSx2076) was identified in *S.agalactiae* <SEQ ID 6095> which encodes the amino acid sequence <SEQ ID 6096>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 30
    >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood =-11.15    Transmembrane    19 - 35 ( 11 - 39)

      ----- Final Results -----
10      bacterial membrane --- Certainty=0.5458 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9911> which encodes amino acid sequence <SEQ ID 9912> was also identified.

15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 6096 (GBS654) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 142 (lane 8 & 10; MW 51.2kDa + lane 9; MW 27kDa). Purified GBS654-GST is shown in Figure 245, lane 11.

20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1968

A DNA sequence (GBSx2077) was identified in *S.agalactiae* <SEQ ID 6097> which encodes the amino acid sequence <SEQ ID 6098>. Analysis of this protein sequence reveals the following:

```

25   Possible site: 14
    >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
30      bacterial cytoplasm --- Certainty=0.4174 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9913> which encodes amino acid sequence <SEQ ID 9914> was also identified.

35 The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:AAF27324 GB:AF178424 unknown [Lactococcus lactis]
      Identities = 26/75 (34%), Positives = 45/75 (59%), Gaps = 4/75 (5%)

    Query: 11 MAFEPKNSSELTQVLKES-LDEEKKEIFSSSEMNIQDFERTKQYQFTLQPSVRKKIDRLSKE 69
40      MAF+ + ++ VL S L + K E+ I E K Y FTL+PSV++ +++L+++
      Sbjct: 1 MAFDVDDKKVKTVLSNSSLAKSKVEL---PKKIESEENKKSYSFTLEPSVKEGLEKLAEK 57

    Query: 70 KGYRSASSFINDFPK 84
      + Y++ S F+ND K
45      Sbjct: 58 QNYKNTSQFLNDLIK 72

```

No corresponding DNA sequence was identified in *S.pyogenes*.

-2218-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics:

Example 1969

5 A DNA sequence (GBSx2078) was identified in *S.agalactiae* <SEQ ID 6099> which encodes the amino acid sequence <SEQ ID 6100>. This protein is predicted to be ParA. Analysis of this protein sequence reveals the following:

Possible site: 45
>>> Seems to have an uncleavable N-term signal seq

10 ----- Final Results -----
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF27325 GB:AF178424 ParA [Lactococcus lactis]
Identities = 49/104 (47%), Positives = 72/104 (69%)

20 Query: 22 L SERLEEFKTEAFDFKTRASYVTAKLFFLGNMIKHNTNSSKELIRSLKNDKSVLAMIPHK 81
L ERL+ FK E D +TR +Y+TA +F+GN I+HNT SS+E + DK +AMIP K
Sbjct: 157 L IERLQNFKDEVIDARTRETYITAIPIYFVGNRI RHNTKSSREFSEKISQDKGTIAMIEK 216

Query: 82 ELFNRSTLDKKSLSYMSDKELYSRDSKFFKEIDFTFRKITDKL 125
ELFNRSTLD L M DK++++ + F++++F F +IT+K+
25 Sbjct: 217 ELFNRSTLDGVPLVEMEKDKDVFNSNKVFYEKLNFAFNEITNKI 260

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 1970

A DNA sequence (GBSx2079) was identified in *S.agalactiae* <SEQ ID 6101> which encodes the amino acid sequence <SEQ ID 6102>. This protein is predicted to be transposase (orfA). Analysis of this protein sequence reveals the following:

35 Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2830(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1971

45 A DNA sequence (GBSx2080) was identified in *S.agalactiae* <SEQ ID 6103> which encodes the amino acid sequence <SEQ ID 6104>. This protein is predicted to be transposase (orfB). Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

-2219-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2618 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB90834 GB:AJ250837 putative transposase [Streptococcus dysgalactiae]
 Identities = 242/259 (93%), Positives = 249/259 (95%)

Query: 1 MCRWLNMPHSSYYYQAVESVSETEFEETIKRIFLDSESRYGSRKIKICLNNEGITLSRRR 60
 MCRWLN+P SSSYYY+AVE VSE E EE+IK IFL+S++RYGSRKIKICLNNEGITLSRRR
 Sbjct: 1 MCRWLNIPRSSYYYKAVEPVSEAELEESIKAIFLESKARYGSRKIKICLNNEGITLSRRR 60

15 Query: 61 IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDRQFKQERPLQALVTDLTYYVRVGNR 120
 IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDRQFK ERPLQALVTDLTYYVRVGNR
 Sbjct: 61 IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDRQFKPERPLQALVTDLTYYVRVGNR 120

20 Query: 121 WAYVCLIIDLYNREIIGLSLGHWHTAELVKQAIQSIPYALTKVKMFHSDRXKEFDNQLID 180
 WAYVCLIIDLYNREIIGLSLGHWHTAELVKQAIQSIPY LTKVKMFHSDR KEF+NQLID
 Sbjct: 121 WAYVCLIIDLYNREIIGLSLGHWHTAELVKQAIQSIPYPLTKVKMFHSDRGKEFNNQLID 180

25 Query: 181 EILEAFGITRSLSQAGCPYDनावेष्ट्याफकीएफवीयूएफटीएफलेलालकटीकडिवह्वनी 240
 EILEAFGITRSLSQAGCPYDनावेष्ट्याफकीएफवीयूएफटीएफलेलालकटी कडिवह्वनी
 Sbjct: 181 EILEAFGITRSLSQAGCPYDनावेष्ट्याफकीएफवीयूएफटीएफलेलालकटीकडिवह्वनी 240

 Query: 241 HRIHGSLNYQTPMTKRLIA 259
 HRIHGSLNYQTPMTKRLIA
 Sbjct: 241 HRIHGSLNYQTPMTKRLIA 259

30

There is also homology to SEQ ID 32.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1972

35 A DNA sequence (GBSx2081) was identified in *S.agalactiae* <SEQ ID 6105> which encodes the amino acid sequence <SEQ ID 6106>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.3325 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1973

50 A DNA sequence (GBSx2082) was identified in *S.agalactiae* <SEQ ID 6107> which encodes the amino acid sequence <SEQ ID 6108>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

-2220-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.4442(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9917> which encodes amino acid sequence <SEQ ID 9918> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAD44095 GB:AF115103 orf359 gp [Streptococcus thermophilus
 bacteriophage Sfi21]
 Identities = 92/357 (25%), Positives = 162/357 (44%), Gaps = 33/357 (9%)

15 Query: 45 RKNQYGKTFETMKEAYDELVRIKYEFANKVSLNENYMTFENYMNKIYLRAYKQK-VQSVT 103
 RK + F T EA ++ + V+++ ++T +Y K + YK+ V +T
 Sbjct: 24 RKPRTKGGFRTKSEAIAAAEMELKLDQNVNVD-DITLYDYF-KQWCEVYKPTVSKIT 81

20 Query: 104 YKTALPHHKLFIQYFGLKPLKAITPRDCEAFRLHIIENYSENYAKNLWSRF-----KACMG 159
 YK + + +FG K LK+IT + + ++ +Y++ +A++ RF KAC+
 Sbjct: 82 YKAYINSQRKIELFFGDKKLKLSITATEYQ---RVLNSYAKTHAQDTVERFNVHVKACIE 137

25 Query: 160 YAERLGYISNMPCKALD---NPRGKHPETPFWTYAEFQTFIKSFDLHDYEELQRFITAIWL 216
 A GYI CK +G+ ET F E++ I ++ + E + A+++
 Sbjct: 138 MAVHEGYIKRNFCKFAKINAKNKGRIETKFLEVEEYERLI--YETSKHPEYASYAALYI 195

30 Query: 217 YYMTGVRVSEGLSLCWEDIDFDKFLKVHTTLEKDENGWYRKDQTKTPAGERLIELDDI 276
 TG+R +E L L +DI D L V+ T + N + TKT + R I LDD
 Sbjct: 196 IAKTGIRFAECLGLTVDIDIKRDTGMLSVNKTWDYKNTGFM---PTKTSSIREIPLDDE 252

35 Query: 277 TIEVLQVWRKNQFANQDITDFIISRFDPFCKSTICRIIKRKAQQVGPVITGKGLRHS 336
 I + +Q D I+ + T+ +I+ R+ + LRH++A
 Sbjct: 253 FINFI-----DQLFPTDDGRILPSLSNNAVNKTLRKIVGRE-----VRVHSLRHTYA 299

Query: 337 SYLINVLKDKDILYVARRMGHADKSTTLNTYSHWFNALDKTVSEEITQNIKSAGLDSI 393
 SYLI D++ V++ +GH + + TL Y+H E+I Q G +++
 Sbjct: 300 SYLI-AHDIDLISVSQVLGHENLNITLEVYAHQLQEOKSRNDEKIKQMWTECGRNAL 355

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6109> which encodes the amino acid sequence <SEQ ID 6110>. Analysis of this protein sequence reveals the following:

40 Possible site: 61
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.5549(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 111/127 (87%), Positives = 119/127 (93%)

50 Query: 242 LKVHTTLEKDENGWYRKDQTKTPAGERLIELDDITIEVLQVWRKNQFANQDITDFIISRF 301
 LKVHTTLEKDENGWYRKDQTKTPAGERLIELDD+TI VL+ WR+NQ N DTDIFIISRF
 Sbjct: 1 LKVHTTLEKDENGWYRKDQTKTPAGERLIELDDVTIVLENWRNQQVNTDTDFIISRF 60

55 Query: 302 GDFPFCKSTICRIIKRKAQQVGPVITGKGLRHSASYLINVLKDKDILYVARRMGHADKST 361
 G+PFCKSTICR+IK KAQ +GVPVITGKGLRHS+ASYLINVLKDKDILYVA+ MGHADKST
 Sbjct: 61 GEPFCKSTICRVIKHKAQSIGVPVITGKGLRHSYASYLINVLKDKDILYVAKCMGHADKST 120

60 Query: 362 TLNTYSH 368
 TLNTYSH
 Sbjct: 121 TLNTYSH 127

-2221-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1974

- 5 A DNA sequence (GBSx2083) was identified in *S.agalactiae* <SEQ ID 6111> which encodes the amino acid sequence <SEQ ID 6112>. Analysis of this protein sequence reveals the following:

Possible site: 20
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3299(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1975

- 20 A DNA sequence (GBSx2084) was identified in *S.agalactiae* <SEQ ID 6113> which encodes the amino acid sequence <SEQ ID 6114>. This protein is predicted to be repressor protein-related protein. Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2721(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 30 A related GBS nucleic acid sequence <SEQ ID 9919> which encodes amino acid sequence <SEQ ID 9920> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:AAC98432 GB:L29324 repressor protein [Streptococcus pneumoniae]
Identities = 38/65 (58%), Positives = 52/65 (79%), Gaps = 1/65 (1%)
Query: 2 MYRRLRDLREDNDFQKYVAEK-LSFTHSAYSKIERGERILSADVIIKLSNLYNVSTDYL 60
M +R+RDLRED+D TQ+YVA+ L+ T SAYSK+E G R++S D +IKL++ YNVS DYL
Sbjct: 1 MLKRIRDLREDDDLTQEYVAKTIILNCTRSAYSKMESGTRLISIDDLIKLADFYNVSLDYL 60
40 Query: 61 LGQTD 65
+G+ D
Sbjct: 61 VGRVD 65

There is also homology to SEQ ID 582.

- 45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2222-

Example 1976

A DNA sequence (GBSx2085) was identified in *S.galactiae* <SEQ ID 6115> which encodes the amino acid sequence <SEQ ID 6116>. This protein is predicted to be relaxase. Analysis of this protein sequence reveals the following:

```

5   Possible site: 40
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.3160(Affirmative) < succ>
10   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:AAC98434 GB:L29324 relaxase [Streptococcus pneumoniae]
    Identities = 223/417 (53%), Positives = 310/417 (73%), Gaps = 5/417 (1%)

    Query: 1   MVITKHYAVHGKKYRRQLIKYILDPKKTNRNLSLISDFGMSNYLDFPDYVELVKMYQNNFL 60
            MVITKH+A+HGK YR +LIKYL+P KT+NL+L+SDFGM NYLDFP Y ELVKMY +NFL
    Sbjct: 1   MVITKHFAIHGKNYRSKLIKYLNPSTKTNLTLSDFGMRNYLDFPSYKELVKMYNDNFL 60

20   Query: 61  SNDQLYDSRFRDQEKQKQKIHAHHIIQSFSPEDKLSPEINRIGYETIKELIGGQYKFIV 120
            SND LY+ R DRQE Q+KIH+HHIIQSFSFSP+D L+PE+INRIGYE KEL GG+++FIV
    Sbjct: 61  SNDTLYEFRHQRQEVNQKRIHSHHIIQSFSPPDHLTPEQINRIGYEAAKELTGGRFRFIV 120

25   Query: 121 ATHVDQDHCHNHIIINSINSQSQKKLKWYALERNLQMISDRISKVAGAKIIPPARYSHR 180
            ATHVD+ H HHNII+NSI+ S KK WDY E NL+M+SDR+SK+AGAKII RYSHR
    Sbjct: 121 ATHVDKGGHIIHNHIIILNSIDQNSDKKFLWDYKAENHNLRMVSDRLSKIAGAKII-ENRYSHR 179

30   Query: 181 DYEYVRRSNHKEYELKQRLFFLMEHSIDFNDFMQAEQLNVKIDFSRKHSRFFMTDRNMKQ 240
            YEYVR++N+KYE+KQR++FL+E+S +F D +KA+ L++KIDF KH +FMTD NMKQ
    Sbjct: 180 QEYVYRKTNYKYEIKQRVYFLIENSKNFEDLKKKAKALHLKIDFRHKHVTYFMTDSNMKQ 239

    Query: 241 VIQGDKLKREPYSKEYFQRYFAKKKIELILEFLLLRNSNFDDLVEKARLLGLELKSXXX 300
            V++ KL++++PY++ YF++ F +++I ILEFLL + + ++L+++A + GL++ K+K
35   Sbjct: 240 VVRDSKLSRKQPNYETFEKKFVQREIINILEFLLPKMKNMNELIQRAEVFGKIIIPKEK 299

    Query: 301 TIDFVLSDGKSCISIPNKSRLKKNLYDTTYFDSYFKEHDVFEVLHNNEVKIEFEKFETQQ 360
            + F DG I + + L K NLY +YF YF + VL N + + + + +
40   Sbjct: 300 HVLFEF-DG---IKLAEQELVKSNSLYSVSYFQDYFNNKNETFVLDNKNLVELYNEEKIIK 355

    Query: 361 LSEILTVBEITEAYETYTKRDAVHEFEVEITEEQIEKIVLDGLFVKVWVGIGQEGEGL 417
            E+ + E + ++Y+ +K RDAVHEFEVE+ QIE++V G+++KV GI ++ L
    Sbjct: 356 EKELPSEBMVWKSQDFKRNDAVHEFEVELNLNQIEEVVEHGIYIKVQFGIDKKDL 412

```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6117> which encodes the amino acid sequence <SEQ ID 6118>. Analysis of this protein sequence reveals the following:

```

    Possible site: 20
    >>> Seems to have no N-terminal signal sequence

50   ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.3114(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

55 An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 103/218 (47%), Positives = 170/218 (77%)

    Query: 393 EEQIEKIVLDGLFVKVWVGIGQEGELIFIPNHQNLNILEQENKKQYQVFIRETSSYFIYHKE 452
            E QIE+++ + +++KV + Q GLIFIPN+QL+I ++EN K+Y+V+IRET+ +FIY+KE
60   Sbjct: 2   EHQIERLIAEDIYIKVSFVSQSGELIFIPNYQLDIRKEENHKYKVIYIRETAQFFIYNKE 61

```


-2223-

Query: 453 DSEMNRFMKGRDLIRQLTFDNKSLPYKRRISLVSLQQKIEEINLLMTLNIQNKSFLBLKD 512
 SE+NR+M+G +LI QLT D+KS+P +RR ++ +L++KIEEI+LL+ L+ +NK + ++KD
 Sbjct: 62 ASELNRYMRGHELICQLTNDKSIPKRRRTIDTLKKKIEEISLLIELDTENKPYQDIKD 121

5 Query: 513 ELVGDIQAQLDIELTNLQDKNTTLNKMAEVVVNLQSDNQDTKQLAKYEC SKMNLSONVTIG 572
 ++V D+AQLD+ +T LQD LNK+AEV++NL +++ + ++LA+Y+ +KMNL+ + I
 Sbjct: 122 DIVKDMAQLDLTITELQDHIAHLNKVAEVLNLNNDIENRRLARYDYAKMNLTAALKIE 181

10 Query: 573 QIESEIEMIQNQLDNKIEEYENAVRKLDEYVRVLNMDK 610
 ++E EIE QN+L+ I+EYE VR+L+++ +L+ K
 Sbjct: 182 EVEKEIETSQNELNISIDEYEVLRRLLEKFGCILSDSK 219

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 1977

A DNA sequence (GBSx2086) was identified in *S.agalactiae* <SEQ ID 6119> which encodes the amino acid sequence <SEQ ID 6120>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4006(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC98436 GB:L29324 unknown [Streptococcus pneumoniae]
 Identities = 53/115 (46%), Positives = 77/115 (66%), Gaps = 2/115 (1%)

30 Query: 5 VREIRKEVNFSIEEYQQIQNFMQEYEQFSPFARGKLLKIDHQPQQLEEWIKYLQHQK 64
 +R IRK+ + E +QI + M ++G + FS F R LL D Q +Q+E+W + QK
 Sbjct: 5 IRSIRKQFRLTETEEKQILDLMREKGGDNFSDFLRKSLLLSDGQ--KQMEKWFNLWKKQK 62

35 Query: 65 VEQIYRDVHEILVLAKLSQSVTMEHLEIILTCIKDLMKIEVITIPLSYSFKDKYM 119
 +EQI RDVHE+ ++AK + VT EH+ I+LTCI++L+KE+E T PLS F +KYM
 Sbjct: 63 LEQISRDVHEVFIIAKTNHQVTHEHVSILLTCIQELIKEVEKTGPLSEDFCNKYM 117

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40

Example 1978

A DNA sequence (GBSx2087) was identified in *S.agalactiae* <SEQ ID 6121> which encodes the amino acid sequence <SEQ ID 6122>. This protein is predicted to be TnpA. Analysis of this protein sequence reveals the following:

45 Possible site: 34
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2935(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC82523 GB:AF027768 TnpA [Serratia marcescens]

-2224-

Identities = 176/413 (42%), Positives = 243/413 (58%), Gaps = 18/413 (4%)

Query: 26 MMFKVEAVGPPERCECGFD-KLYKHSSRNQLIMDLPIRLKRVGLHLNRRRYKCRECGST 84
 M F+V+ V P C ECG + + R+ DLPI KRV L + RRRY CR C +T
 5 Sbjct: 1 MHFQVD-VPDPIACEECGVQGEFVRFGKRDVPYRDLPIHGKRVTLWVRRRYTCRACKTT 59

Query: 85 IS-----VDEKRSMTKRLLSIQEQSMSTFVEVAESVGVDEKTI RNVFKDYVALKERE 138
 VD R MT RL + +++S + + VA G+DEKT+R++F R
 10 Sbjct: 60 FRPQLPEMVDGFR-MTLRLHEYVEKESFNHPYTFVAAQTGLDEKTVRDIFNARAEFLGRW 118

Query: 139 YQFETPKWL GIDEIHIIRRPRLVLTNIERRTIYDIKPNRNKETVIQRLSEISDRTYIEYV 198
 ++FETP+ LGIDE+++ +R R +LTNIE RT+ D+ R ++ V L ++ DR +E V
 15 Sbjct: 119 HRFETPRILGIDELYLNKRYRCILTNIEERTLLDLLLATRRQDVVINYLMKLKDRQKVEIV 178

Query: 199 TMDMWKPYKDAVNTILPQAKVVVDKFHVVRMANQALDNVRKSLKAHMSQKERRTLMRERF 258
 +MDMW PY+ AV +LPQA++VVDKFHVVRMAN AL+ VRK L+ + + RTL +R
 20 Sbjct: 179 SMDMWNPYRAAVKAVLPQARIVVDKFHVVRMANDALERVRKGLRKELKPSQSRTLKGDRK 238

Query: 259 ILLKRKHDLNERESFLDITWLGNLPALKEAYELKEEFYWIWDTDPDEGHLRYSQWRHRC 318
 ILLKR H++++RE +++TW G P L AYE KE FY IWD + +W
 25 Sbjct: 239 ILLKRAHEVSDRERLIMETWTGAFQQLAAYEHKERFYGIWDATTRLQAEALDEW-IAT 297

Query: 319 MSSNSKDAYKDLVRAVDNWHVEIFNYF--DKRLTNAYTESINSIIRQVERMGRGYSFDAL 376
 + K+ + DLVRAN NW E YF D +TNAYTESIN + + R GRGYSF+ +
 30 Sbjct: 298 IPKGQKEVWSDLVRAVGNWRBETMTYFETDMPVTNAYTESINRLAKDKNREGRGYSFEVM 357

Query: 377 RAKILFNEKLHKRKRPRFNSSAFNKAMLYDTFNWYEVNDHDITDNLGVDFSTL 429
 RA++L+ K HKK+ P S F K + Y + D N GVD ST+
 35 Sbjct: 358 RARMLYTTK-HKKKAPTAKVSPFYKTTI-----GYGLPDFAEELNYGVLDLSTI 404

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1979

35 A DNA sequence (GBSx2088) was identified in *S.agalactiae* <SEQ ID 6123> which encodes the amino acid sequence <SEQ ID 6124>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2115(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA70224 GB:Y09024 mercuric reductase [Bacillus cereus]

Identities = 412/546 (75%), Positives = 484/546 (88%)

50 Query: 1 MNKFKVNIISGMTCTGCEKHIVESALEKIGAKNISSYRRGEAVFELPDIEVESAIKAIDE 60
 M K++V++ GMTCTGCE+HV ALE +GA IE +RRGEAVFELP+ + VE+A KAI +
 Sbjct: 1 MKKYRVDVQGMTCTGCEEHVAVALENMGATGIEVDFFRGEAVFELPNALGVETAKKAISD 60

55 Query: 61 ANYQAGEIEEVSSLENVALINEDNYDLLIIGSGAAAFSSAIKAIYGAQVGMIERGTGVGG 120
 A YQ G+ EEV S E V L NE +YD +IIGSG AAFSSAI+A++YGAKV MIERGT+GG
 Sbjct: 61 AKYQPGKAEVQSQEMVQLGNEGDYDYIIIGSGGAAAFSSAIEAVKYGAQVAMIERGTIGG 120

Query: 121 TCVNIGCVPSKTLRAGEINHLKSDNPFIFGLQTSAGEVDLASLITQDKLVSELNQQYM 180
 TCVNIGCVPSKTLRAGEINHL+K+NPFF+GL TSAGEVDLA LI QK++LV+ELRN KY+
 60 Sbjct: 121 TCVNIGCVPSKTLRAGEINHLAKNPNFVGLHTSAGEVDLAPLIKQKNELVTELNRNSKYV 180

-2225-

5 Query: 181 DLIDEYNFDLIKGEAKFVDASTVEVNGTKLSAKRFLIATGASPSLPQISGLEKMDYLTST 240
 DLID+Y F+LI+GEAKFVD TVEVNG +SAKRFLIATGASP+ P I GL ++DYLTST
 Sbjct: 181 DLIDDYGFELIEGEAKFVDEKTVENVGAPISAKRFLIATGASPAKPNIPGLNEVDYLTST 240

10 Query: 241 TLLELKKIPKRLTVIGSGYIGMELGQLFHHLGSEITLMQSRERLLKEYDPEISESVEKAL 300
 +LLELKK+PKRL VIGSGYIGMELGQLFH+LGSE+TL+QSRERLLKEYDPEISESVEK+L
 Sbjct: 241 SLLELKKVPKRLVVIGSGYIGMELGQLFHNLGSEVTLIQSRERLLKEYDPEISESVEKSL 300

15 Query: 301 IEQGINLVKGATFERVEQSGEIKRVYTVNGSREVIESDQLLVATGRKENTDSNLNSAAG 360
 +EQGINLVKGAT+ER+EQ+G+IK+V+V VNG + +IE+DQLLVATGR PNT +LNL AAG
 Sbjct: 301 VEQGINLVKGATYERIEQNGDIKKVHVEVNGKIRIIEADQLLVATGRTPNTATNLNRAAG 360

20 Query: 421 VVPAVTFNTPTVATVGLTEEQAKEKGYDVKTSVLPLDAVPRAIVNRETTGVFKLVADAET 480
 VVP VTFT P +ATVGLTE+QAKE GY+VKTSVLPLDAVPRA+VNRETTGVFKLVAD++T
 Sbjct: 421 VVPGVTFPTAPAIATVGLTEQAKENGVEKTSVLPLDAVPRALVNRETTGVFKLVADSKT 480

25 Query: 481 LKVLGVHIVSENAGDVIYAASLAVKFGLTIEDLTETLAPYLTMAEGLKLVALTFDKDISK 540
 +KVLG H+V+ENAGDVIYAA+LAVKFGLT++D+ ETLAPYLTMAEGLKL ALTFDKDISK
 Sbjct: 481 MKVLGAHVVAENAGDVIYAATLAVKFGLTVDDIRETLAPYLTMAEGLKLAALTFDKDISK 540

Query: 541 LSCCAG 546
 LSCCAG
 Sbjct: 541 LSCCAG 546

30 There is also homology to SEQ ID 1820.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1980

35 A DNA sequence (GBSx2089) was identified in *S.galactiae* <SEQ ID 6125> which encodes the amino acid sequence <SEQ ID 6126>. This protein is predicted to be regulatory protein. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4529(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA83973 GB:AF138877 mercury resistance operon negative
 regulator MerR1 [Bacillus sp. RC607]
 Identities = 83/129 (64%), Positives = 104/129 (80%)

50 Query: 1 MIYRISEFADKCGVNKETIRYERKNLLQEPHRTGAGYRIYSYDDVKRVGFIKRIQEFGF 60
 M +RI E ADKCGVNKETIRYER L+ EP RTE GYR+YS V R+ FIKR+QE GF
 Sbjct: 1 MKFRIGELADKCGVNKETIRYERLGLIPEPERTEKGYRMYSQQTVDRLHFIKRMQELGF 60

55 Query: 61 SLSEIYKLLGVVDKDEVRCQDMFEFVSKKQKEVQKQIEDLKRIETMLDDLKQRCPEDEK 120
 +L+EI KLLGVVD+DE +C+DM++F K +++Q++IEDLKRIE ML DLK+RCP+ K +
 Sbjct: 61 TLNEIDKLLGVVDREAKCRDMYDFITLKIEDIQRKIEDLKRIERMLMDLKERCPCENKDI 120

Query: 121 HSCPIIETL 129
 + CPIIETL
 60 Sbjct: 121 YECPIIETL 129

-2226-

There is also homology to SEQ ID 1712.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 1981

A DNA sequence (GBSx2090) was identified in *S.agalactiae* <SEQ ID 6127> which encodes the amino acid sequence <SEQ ID 6128>. Analysis of this protein sequence reveals the following:

```

Possible site: 44
>>> Seems to have no N-terminal signal sequence
10    INTEGRAL    Likelihood = -7.86    Transmembrane    80 - 96 ( 78 - 100)

----- Final Results -----
        bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8943> and protein <SEQ ID 8944> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 8
McG: Discrim Score:    -13.52
25    GvH: Signal Score (-7.5): -6.14
        Possible site: 44
>>> Seems to have no N-terminal signal sequence
ALOM program    count: 1 value: -7.86 threshold: 0.0
        INTEGRAL    Likelihood = -7.86    Transmembrane    80 - 96 ( 78 - 100)
30    PERIPHERAL Likelihood = 1.80    136
        modified ALOM score:    2.07

*** Reasoning Step: 3

35    ----- Final Results -----
        bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

40 The protein has homology with the following sequences in the databases:

```

ORF02021(439 - 666 of 1080)
GP|451734|gb|AAA18975.1||U05143(9 - 46 of 46) envelope glycoprotein {Simian
immunodeficiency virus} GP|451744|gb|AAA18980.1||U05148 envelope glycoprotein {Simian
45 immunodeficiency virus}
%Match = 3.2
%Identity = 38.5 %Similarity = 64.1
Matches = 15 Mismatches = 13 Conservative Sub.s = 10

336      366      396      426      456      486      516      546
50    RIPVQFKGCCDDYYNENVGYPLSRINLEHYLTGGVLYFVVYSKDVSPSTVTYASLTPKVIKNVLPASDKKKRIKKEDIFL
        :|| | : ||:|:|:|: |
        WGLTGNAGTTPTATTTTTTPRVVENVINESN-----
        10          20          30

```

-2227-

```

576      606      636      666      696      726      756      786
LFWMAIIAKLLILPYPALQTSYKSRPCLRRSSLRKLTQIPFSIVTKVGNINMKSITAFLOVKAYILPCLAKGPARIMV*W
      ||:::|  |||
-----PCIKDNSCAGLEQEP
                        40

```

SEQ ID 8944 (GBS415) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 3; MW 21.2kDa).

Example 1982

A DNA sequence (GBSx2092) was identified in *S. agalactiae* <SEQ ID 6129> which encodes the amino acid sequence <SEQ ID 6130>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have no N-terminal signal sequence

```

```

----- Final Results -----

```

```

bacterial cytoplasm --- Certainty=0.3402(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1983

A DNA sequence (GBSx2093) was identified in *S. agalactiae* <SEQ ID 6131> which encodes the amino acid sequence <SEQ ID 6132>. This protein is predicted to be ATPase. Analysis of this protein sequence reveals the following:

```

Possible site: 27
>>> Seems to have no N-terminal signal sequence

```

```

INTEGRAL Likelihood = -10.08 Transmembrane 324 - 340 ( 317 - 343)
INTEGRAL Likelihood = -5.73 Transmembrane 662 - 678 ( 660 - 690)
INTEGRAL Likelihood = -5.41 Transmembrane 350 - 366 ( 346 - 378)
INTEGRAL Likelihood = -3.40 Transmembrane 94 - 110 ( 93 - 110)
INTEGRAL Likelihood = -2.87 Transmembrane 681 - 697 ( 680 - 699)
INTEGRAL Likelihood = -1.38 Transmembrane 148 - 164 ( 148 - 164)

```

```

----- Final Results -----

```

```

bacterial membrane --- Certainty=0.5034(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAA22858 GB:M90750 cadmium-efflux ATPase [Bacillus firmus]
Identities = 486/725 (67%), Positives = 584/725 (80%), Gaps = 18/725 (2%)

```

```

Query: 1 MSRGKAKQSEKEMKAYRVQGFTCTNCAAI FENN VKELPGVQDAKVNFGASKVYVKGTTTI 60
MS KA SE+EMKAYRVQGFTC NCA FE NVK+L GV+DAKVNFGASK+ V G TI
Sbjct: 1 MSDQKAITSEQEMKAYRVQGFTCANCAGKFEKNVKQLSGVEDAKVNFGASKIAVYGNATI 60

```

```

Query: 61 EELEKAGAFENLKIRDEKEQRVGGE-----PFWKQKENIKVYISALLLVSWFL 109
EELEKAGAFENLK+ EK R + PF+K K + +Y S LL+ +
Sbjct: 61 EELEKAGAFENLKVTPEKSARQASQEVKEDTKEDKVPFYK-KHSTLLYAS-LLITFGYLS 118

```

```

Query: 110 GEQYGEHVLPTIGYAASILIGGYSLFIKGLKNLRLNFDNMNTLMTIAIGAAIIGEWGE 169
GEE+++ T+ + AS+ IGG SLF GL+NL R FDM TLMT+A+IG AIIGEW E

```

-2228-

Sbjct: 119 SYVNGEENIVTTLFLASMFIGGLSLFKVGLQNLRRFEDMKTLMTVAVIGGAIIGEWAE 178

Query: 170 GATVVILFAISEALERYSMCDKARQSIESLMDIAPKEALIRRGNEEMMIHVDEIQVGDIMI 229
A VVILFAISEALER+SMD+ARQSI SLMDIAPKEAL++R +E+MIHVD+I VGDIMI

5 Sbjct: 179 VAIIVVILFAISEALERFSMDRARQSIERSLMDIAPKEALVKRNGQEIMIHVDDIAVGDIMI 238

Query: 230 VKPGQKLAMDGIIVVKGSTSLNQAAITGESVPVTKITNDEVFAGTLNEEGLLEVKTVE 289
VKPGQK+AMDG+VV G S +NQ AITGESVPV K ++EVFAGTLNEEGLLEV++TK VE

10 Sbjct: 239 VKPGQKIAMDGVVVSQSAVNTAITGESVPVEKTVDNFVAGTLNEEGLLEVEITKLVE 298

Query: 290 DTTLSKIIHLVEEAQAERAPSQAQFVDKFAKYYTPAIVILALLIAVVPPL-FGGDWSQWYI 348
DTT+SKIIHLVEEAQ ERAPSQAQFVDKFAKYYTP I+I+A L+A+VPPL F G W WIY

15 Sbjct: 299 DTTISKIIHLVEEAQGERAPSQAQFVDKFAKYYTPIIMIIATLVAIVPPLFFDGSWETWIY 358

Query: 349 QGLAVLVVGCPCALVSTPFAVAVTAIGNAAKNGVLIKGGIHLAAGHLKAIADFDTGLT 408
QGLAVLVVGCPCALV+STP+++V+AIGNAAK GVL+KGG++LE G LKAIADFDTGLT

20 Sbjct: 359 QGLAVLVVGCPCALVISTPISIVSAIGNAAKGVLVKGGVYLEMGALKAIAFDKTGLT 418

Query: 409 KGIPAVTD--IVTYGRNENELITITSAIEKGSQHPLASAIMKAEENGLKFNEVTVEDFQ 466
KG+PAVTD ++ NE EL++I +A+E SQHPLASAIM+KAE + +++V VEDF

25 Sbjct: 419 KGVPAVTDYNVLNKQINEKELLSIITALEYRSQHPLASAIMKAEENITYSDVQVEDFS 478

Query: 467 SITGKGVKAKINNEYVYVGSQNLFEELHGSISSDKKEKIADMQTQGTVMVLGTEKEIL 525
SITGKG+K +N YY+GS LF+E L D ++ + +Q QGKT M++GTEKEIL

30 Sbjct: 479 SITGKGIKGI VNGTTYIIGSPKLFKELLTNDFDKLEQNVTTLQNGKTAMIIGTEKEIL 538

Query: 526 SFIADADEMRESSKEVIGKLNNGI-ETVMLTGDNQTATAIGKQVGVSDIKADLLPEDK 584
+ IADADE+RESSKE++ KL+ +GI +T+MLTGDN+ TA AIG QVGVSDI+A+L+P+DK

35 Sbjct: 539 AVIADADEVRESSKEILQKLHQLGIKKTIMLTGDNKGTAIGAIGQVGVSDIEALMPQDK 598

Query: 585 LNFIKELREKHQSVGMVGDGVNDAPALAASTVGVAMGGAGTDTALETADIALMSDDLK 644
L+FIK+LR ++ +V MVDGVDNDAPALAASTVG+AMGGAGTDTALETAD+ALM DDL KL

40 Sbjct: 599 LDFIKQLRSEYGNVAMVGDGVNDAPALAASTVGIAMGGAGTDTALETADVALMGDDLRL 658

Query: 645 PYTIKLSRKALAIKQNTIFSLAIKLVALLVMPGWLTLWIAIFADMGATLLVTLNLSRL 704
P T+KLSRK L IIK NITF++AIK +A LLV+PGWLTLWIAI +DMGATLLV LN LRL

45 Sbjct: 659 PSTVKLSRKTLNIIKANITFAIAIKFIASLLVPGWLTLWIAILSDMGATLLVALNGLRL 718

Query: 705 LKIKE 709
+++KE

50 Sbjct: 719 MRVKE 723

There is also homology to SEQ ID 3506.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1984

A DNA sequence (GBSx2094) was identified in *S.agalactiae* <SEQ ID 6133> which encodes the amino acid sequence <SEQ ID 6134>. Analysis of this protein sequence reveals the following:

Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0779(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2229-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1985

A DNA sequence (GBSx2095) was identified in *S.agalactiae* <SEQ ID 6135> which encodes the amino acid sequence <SEQ ID 6136>. Analysis of this protein sequence reveals the following:

```

Possible site: 58
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -8.92    Transmembrane    123 - 139 ( 115 - 145)
    INTEGRAL    Likelihood = -6.74    Transmembrane    172 - 188 ( 167 - 190)
10    INTEGRAL    Likelihood = -1.81    Transmembrane    80 - 96 ( 80 - 96)

----- Final Results -----
                bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9923> which encodes amino acid sequence <SEQ ID 9924> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 4216.

A related GBS gene <SEQ ID 8945> and protein <SEQ ID 8946> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 10
McG: Discrim Score:      -6.41
25    GvH: Signal Score (-7.5): -2.23
    Possible site: 58
    >>> Seems to have no N-terminal signal sequence
    ALOM program    count: 3 value: -8.92 threshold: 0.0
    INTEGRAL    Likelihood = -8.92    Transmembrane    123 - 139 ( 115 - 145)
30    INTEGRAL    Likelihood = -6.74    Transmembrane    172 - 188 ( 167 - 190)
    INTEGRAL    Likelihood = -1.81    Transmembrane    80 - 96 ( 80 - 96)
    PERIPHERAL    Likelihood = 2.92      46
    modified ALOM score: 2.28

35    *** Reasoning Step: 3

----- Final Results -----
                bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
40    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1986

A DNA sequence (GBSx2096) was identified in *S.agalactiae* <SEQ ID 6137> which encodes the amino acid sequence <SEQ ID 6138>. This protein is predicted to be histidine rich P type ATPase (HRA-1) (copB). Analysis of this protein sequence reveals the following:

```

Possible site: 54
>>> Seems to have no N-terminal signal sequence
50    INTEGRAL    Likelihood = -13.37    Transmembrane    318 - 334 ( 307 - 345)
    INTEGRAL    Likelihood = -5.84    Transmembrane    347 - 363 ( 335 - 364)
    INTEGRAL    Likelihood = -5.15    Transmembrane    88 - 104 ( 86 - 112)

```

-2230-

INTEGRAL Likelihood = -5.04 Transmembrane 651 - 667 (649 - 669)
 INTEGRAL Likelihood = -4.30 Transmembrane 156 - 172 (155 - 173)
 INTEGRAL Likelihood = -4.30 Transmembrane 669 - 685 (668 - 690)
 INTEGRAL Likelihood = -3.03 Transmembrane 62 - 78 (60 - 80)

----- Final Results -----

bacterial membrane --- Certainty=0.6349(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA62113 GB:U16658 histidine rich P type ATPase [Escherichia coli]

Identities = 598/731 (81%), Positives = 651/731 (88%), Gaps = 36/731 (4%)

Query: 1 MRNNKHHSSSHHHNHGDDIDHSHKHDHNEHESQMDHS----- 36
 MRNNK+HSSSHHHNHGD++HSHKHDHNEHESQMDHS
 Sbjct: 1 MRNNKQHSSSHHHNHGDMHESKHDHNEHESQMDHSAMGHGCMGGHAHHHHGDMHSHKHD 60

Query: 37 -----NMDHSEMDHGAMGGHAHHHHGHSFKEIFLKSPLGLIAILLITPMMDIQL 84
 MD+SEMDHGAMGGHAHHHHGHSFK+IFLKSPLGLIAILLITP+M IQL
 Sbjct: 61 HNEMKHESQMDHSHKMDYSEMDHGAMGGHAHHHHGHSFKDIFLKSPLGLIAILLITPLMGIQL 120

Query: 85 PFQIIFPYADVVAVLATILYIFGGKPFYMGAKDEFNSKAPGMMSLITLGITVSYAYSIV 144
 PFQIIFPYADVVAVLATILYIFGGKPF MGAKDEFNSK PGMMSLITLGITVSYAYSIV
 Sbjct: 121 PFQIIFPYADVVAVLATILYIFGGKPFMLGAKDEFNSKVPGMMSLITLGITVSYAYSIV 180

Query: 145 AVAARYVTGEHVMDFFEFTTLILIMLLGHWIEMKALGEAGDAQKALAEVLVPKDAHVVLE 204
 AVAARYVTGE VMDFFFEFTTLILIMLLGHWIEMKALGEAG+AQKALAEVLVPKDAHVVLE
 Sbjct: 181 AVAARYVTGEPVMDFFFEFTTLILIMLLGHWIEMKALGEAGNAQKALAEVLVPKDAHVVLE 240

Query: 205 DDSIETRPVSELQIGDVIRVQAGENVPADGIIIRGESRVNEALVTGESKPIEKKTGDEVI 264
 DDSIETRPV++LQ+GD+IRVQAGENVPADG I RGESRVNEALVTGESKPIEK GDEVI
 Sbjct: 241 DDSIETRPVADLQVGDLIRVQAGENVPADGTIORGESRVNEALVTGESKPIEKNPGDEVI 300

Query: 265 GGSNTGGGVLYVEIKQTGDQSFISQVQTLISQAQSQPSRAENVAQKVASWLFYIAVVVAL 324
 GGSNTG GVLVVEIKQTGD+SFISQVQTLISQAQSQPSRAEN+AQKVA WLFYIAV+ AL
 Sbjct: 301 GGSNTGDGVLYVEIKQTGDKSFISQVQTLISQAQSQPSRAENLAQKVAGWLFYIAVIAAL 360

Query: 325 IALLIWTIIADLPTAVIFTVTALVIACPHALGLAIPLVSRSTSLGASRGLLVKNREALE 384
 IAL+IW +IAD+PTAVIFTVT LVIACPHALGLAIPLV +RSTSLGASRGLLVK+R+ALE
 Sbjct: 361 IALVIWVIADVPTAVIFTVTTLVIACPHALGLAIPLVTRSTSLGASRGLLVKDRDALE 420

Query: 385 LTTKADVMVLDTGTTLTGFEKVLVDVTVLSDKYSEEEITGLLAGIEAGSSHPIAQSIIVNH 444
 LTT ADVMVLDTGTTLTGFEKVLVDV + +DKY+++EI LL+GIE GSSHPIAQSI+++
 Sbjct: 421 LTTNADVMVLDTGTTLTGFEKVLVDVELFNDKYTKDEIVALLSGIEGGSSHPIAQSIISY 480

Query: 445 AEAKGIKSVSFDSEIVSGAGIEGEANGHHYQLISQKAYGKALRMDIPKGATLSILVENN 504
 AE +GI+ VSFDSI+++SGAG+EG+ANGH YQLISQKAYG+ L MDIPKGAT+S+LVEN+
 Sbjct: 481 AEQQGIRPVSFDSIDVMSGAGVEGQANGHRYQLISQKAYGRNLDMDIPKGATISVLVEND 540

Query: 505 EAIGAVALGDELKETSRLNLEVLKKYGTIEPLMATGDNEEAQGVAEVLGIQYQANQSPED 564
 EAIGAVALGDELK TS++LI+ LKK I+P+MATGDNE+AAQG AE+LGI Y ANQSP+D
 Sbjct: 541 EAIGAVALGDELKPTSKDLIQAQKKNKIPIMATGDNEKAAQGAEEILGIDYLANQSPQD 600

Query: 565 KYKLVESMKNQNKTVIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQSDPGDI 624
 KY+LVE +K + K VIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQ PGDI
 Sbjct: 601 KYELVEKLKAEKKVIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQYSPGDI 660

Query: 625 ESFIELANKTTRKMKQNLVWGAGYNFIAIPIAAGLLAPIGITLGPAGAVLMSLSTVIVA 684
 SFIELA KTRKMK+NLVWGAGYNFIAIPIAAG+LAPIGITL PA AVLMSLSTVIVA
 Sbjct: 661 ASFIELAQKTRKMKENLVWGAGYNFIAIPIAAGILAPIGITLSPAAVAVLMSLSTVIVA 720

Query: 685 INAMTLKLEPK 695
 INAMTLKLEPK
 Sbjct: 721 INAMTLKLEPK 731

-2231-

There is also homology to SEQ ID 3506.

A related GBS gene <SEQ ID 8947> and protein <SEQ ID 8948> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1      Crend: 7
5  McG: Discrim Score:      -19.12
   GvH: Signal Score (-7.5): -3.71
      Possible site: 27
   >>> Seems to have no N-terminal signal sequence
ALOM program   count: 7 value: -13.37 threshold: 0.0
10  INTEGRAL    Likelihood ==-13.37  Transmembrane 291 - 307 ( 280 - 318)
     INTEGRAL    Likelihood = -5.84   Transmembrane 320 - 336 ( 308 - 337)
     INTEGRAL    Likelihood = -5.15   Transmembrane 61 - 77 ( 59 - 85)
     INTEGRAL    Likelihood = -5.04   Transmembrane 624 - 640 ( 622 - 642)
15  INTEGRAL    Likelihood = -4.30   Transmembrane 129 - 145 ( 128 - 146)
     INTEGRAL    Likelihood = -4.30   Transmembrane 642 - 658 ( 641 - 663)
     INTEGRAL    Likelihood = -3.03   Transmembrane 35 - 51 ( 33 - 53)
     PERIPHERAL  Likelihood = 0.74    103
   modified ALOM score: 3.17

20  *** Reasoning Step: 3

     ----- Final Results -----
           bacterial membrane --- Certainty=0.6349(Affirmative) < succ>
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

[illegible]

[illegible]

35 Example 1987

40 Possible site: 59
 >>> Seems to have no N-terminal signal sequence

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:CAA40599 GB:X57326 ORF-1 [Thiobacillus ferrooxidans]
Identities = 26/65 (40%), Positives = 40/65 (61%), Gaps = 2/65 (3%)

Query: 1 MKQEILL--DGVKCAGCANTVQERFSAIEGVESVEVDLATKKAVLESQTEIDTETLNAAL 58
M Q+I L G+ CA CA++V++ I G++S +V L+T +A + Q+ I TE L AA+
Sbjct: 1 MSQKIFLRITGMTCAHCAHSVEKALLGIHGIDSAQVSLATNQAQEVFLQSSIPTEALLAAV 60

55 Query: 59 AETNY 63
+ Y
Sbjct: 61 TOAGY 65

There is also homology to SEQ ID 3510.

-2233-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1988

A DNA sequence (GBSx2098) was identified in *S.agalactiae* <SEQ ID 6141> which encodes the amino acid sequence <SEQ ID 6142>. Analysis of this protein sequence reveals the following:

Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.3220(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1989

A DNA sequence (GBSx2099) was identified in *S.agalactiae* <SEQ ID 6143> which encodes the amino acid sequence <SEQ ID 6144>. This protein is predicted to be heavy-metal transporting P-type ATPase (b0484). Analysis of this protein sequence reveals the following:

Possible site: 27
>>> Seems to have no N-terminal signal sequence
25 INTEGRAL Likelihood = -4.09 Transmembrane 131 - 147 (130 - 150)

----- Final Results -----
 bacterial membrane --- Certainty=0.2635(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB01764 GB:U42410 heavy-metal transporting P-type ATPase
 [Proteus mirabilis]
Identities = 98/153 (64%), Positives = 123/153 (80%)
35 Query: 2 KAVKALRRRGVEVIMITGDNKRTAKAIAKQVGIDSVLSEVLPEDKAEVKKLQEAGKKVA 61
 +A+KAL G++V MITGDNK TAKAIAKQ+GID +++EVL+P+ K +K+L + G KVA
 Sbjct: 649 EAIKALHALGLKVAMITGDNKATAKAIKQIGIDEIVAELPDGKVAALKQLSQKGDKVA 708
40 Query: 62 MVGDGINDAPALAQANVGIAGVSGTDVAIESADIVLMRNDLTAVLTITIDLSHATLRNIQ 121
 VGDGINDAPALAQ+VG+A+G+GTDVAIE+AD+VLM DL V+ I LS AT+RNKIQ
 Sbjct: 709 FVGDGINDAPALAQADVGLAIGTGTDAIEAADVVLMSGDLRGVVDALIALSQATIRNIQ 768
 Query: 122 NLFWAFAYNLVGIPVAMGLLYIFGGLMSPMLA 154
45 NLFW FAYN + IPVA G+LY G+L+SP+ A
 Sbjct: 769 NLFWTFAYNALLIPVAAGMLYPINGMLLSPIFA 801

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3505> which encodes the amino acid sequence <SEQ ID 3506>. Analysis of this protein sequence reveals the following:

50 Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -10.83 Transmembrane 328 - 344 (314 - 348)

-2234-

5
 INTEGRAL Likelihood = -7.01 Transmembrane 354 - 370 (347 - 377)
 INTEGRAL Likelihood = -3.24 Transmembrane 101 - 117 (100 - 117)
 INTEGRAL Likelihood = -2.97 Transmembrane 165 - 181 (165 - 185)
 INTEGRAL Likelihood = -2.34 Transmembrane 665 - 681 (662 - 684)
 INTEGRAL Likelihood = -2.18 Transmembrane 67 - 83 (66 - 83)
 INTEGRAL Likelihood = -0.64 Transmembrane 491 - 507 (490 - 508)
 INTEGRAL Likelihood = -0.59 Transmembrane 691 - 707 (691 - 707)
 INTEGRAL Likelihood = -0.43 Transmembrane 140 - 156 (139 - 156)

10
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15 An alignment of the GAS and GBS proteins is shown below.

Identities = 92/152 (60%), Positives = 123/152 (80%)

20
 Query: 4 VKALRRRGVEVIMITGDNKRTAKAIAKQVGIDSVLSEVLPEDKAEVKKLQEAGKKVAMV 63
 V+AL + G+ IM+TGD+ TAKAIA QVGI V+S+VLP+ KA + L+ G+KVAMV
 Sbjet: 544 VEALHQLGIHTIMLTGDHDATAKAIASQVGITDVISQVLPDQKAGVIADLRSQGRKVAMV 603

 Query: 64 GDGINDAPALAQANVGIAVSGSTDVAIESADIVLMRNDLTAVLTITIDLSHATLRNIQNL 123
 GDGINDAPALA A++GIA+GSGTD+AIESAD++IM+ D+ ++ + LS T+R +K+NL
 Sbjet: 604 GDGINDAPALAVADIGIAMSGSTDIAIESADVILMKPMDLDLVKAMSLSRVTMRIVKENL 663

 25
 Query: 124 FWAFAYNLVGIPVAMGLLYIFGGLLMSPMLAG 155
 FWAF YN++ IPVAMGLL++FGG L++PMLAG
 Sbjet: 664 FWAFIYNVLMIPVAMGLLHLFGGPLLNPMMLAG 695

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1990

A DNA sequence (GBSx2100) was identified in *S.agalactiae* <SEQ ID 6145> which encodes the amino acid sequence <SEQ ID 6146>. This protein is predicted to be CopY. Analysis of this protein sequence
 35 reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

40
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2067(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45
 >GP:AAG10085 GB:AF296446 CopY [Streptococcus mutans]
 Identities = 63/139 (45%), Positives = 96/139 (68%)

 Query: 8 TSITDAEWEVMRVVWANDLVTSKTVISVLKEKMDWTESTIKTILGRIVEKGVLTNTBQEGR 67
 TSI++AEWEVMRVVWA + +S +I++L W+ STIKT++ RL EKG L ++++GR
 50 Sbjet: 2 TSISNAEWEVMRVVWAKQMTSSSEIIAILSRITYCWSASTIKTLITRLSEKGYLTSQRQGR 61

 Query: 68 KFIYTANIVEKEAVRDFIEDIFNRICKKKVGNVIGSIIEDHVLVSFDDIDRLEKILEIKKS 127
 K+IY++ I E+EA+ ++F+RIC K +I ++E+ ++ DI++LE +L KK+
 Sbjet: 62 KYIYSSLISEEEALEQQVSEVFSRICVTKHQALIRHLVEETPMTLSDIEKLEALLLSKKA 121

 55
 Query: 128 FAVEEVDCQCTEGQCDCHE 146
 AV EV C C GQC C+E
 Sbjet: 122 NAVPEVKNCIVGQCSCYE 140

60 There is also homology to SEQ ID 3502.

-2235-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1991

A DNA sequence (GBSx2101) was identified in *S.agalactiae* <SEQ ID 6147> which encodes the amino acid sequence <SEQ ID 6148>. Analysis of this protein sequence reveals the following:

Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.2829(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1992

A DNA sequence (GBSx2102) was identified in *S.agalactiae* <SEQ ID 6149> which encodes the amino acid sequence <SEQ ID 6150>. This protein is predicted to be DS RF protein. Analysis of this protein sequence reveals the following:

Possible site: 57
>>> Seems to have a cleavable N-term signal seq.
25 INTEGRAL Likelihood = -13.21 Transmembrane 142 - 158 (136 - 169)
 INTEGRAL Likelihood = -3.45 Transmembrane 70 - 86 (66 - 88)
 INTEGRAL Likelihood = -3.13 Transmembrane 178 - 194 (176 - 195)

----- Final Results -----
30 bacterial membrane --- Certainty=0.6286(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:AAA26611 GB:L10909 putative [Staphylococcus aureus]
 Identities = 98/204 (48%), Positives = 148/204 (72%), Gaps = 3/204 (1%)

Query: 4 TIISAIGVYISTSIDYLVLIILFAQLSQNKQKWHIYAGQYLGTLGVGASLVAAY-VVN 62
 TI++A VY++T IDYL++LI+LF+Q+ + + K HI+ GQY+GT +++GASL+ A VVN
40 Sbjct: 18 TILTATAVYVATGIDYLVILILLFSQVKKQVK-HIWIGQYIGTAIVIGASLLVAQGVVN 76

Query: 63 FVPEAWMVGLLGLIPIYLGIRFAIVGEGEHEEEEEIIEERLEQSKANQLFWITVTLTIASG 122
 +P+ W++GLLGL+P+YLG++ I GE E+E+E I+ K NQLF T+ + +AS
Sbjct: 77 LIPQQWVIGLLGLLPLYLGVKIWKGE-EDDESSILSLFSSGKFNQLFLIMIFIVLASS 135

45 Query: 123 GDNLGIYIPYFASLDWSQTLVVLVFAIGIIFCELSWVLSSIPLISETIEKYQRIIVPL 182
 D+ IYIPYF +L S+ +V +VF I + + C +S+ L+S ISETIEKY+R IVP+
Sbjct: 136 ADDFSIYIPYFTTSLMSSEIFIVTIVFLIMVGVLCYVSYRLASPDFISETIEKYERWIVPI 195

50 Query: 183 VFIPLGLYIMYESGTIETFLNFIL 206
 VFI LG+YI++E+GT ++F+L
Sbjct: 196 VFICLGIYILFENGTSNALISFLL 219

-2236-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6151> which encodes the amino acid sequence <SEQ ID 6152>. Analysis of this protein sequence reveals the following:

```

Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
5   INTEGRAL    Likelihood = -13.16   Transmembrane  143 - 159 ( 135 - 165)
   INTEGRAL    Likelihood = -9.13    Transmembrane   49 - 65 (  43 - 71)
   INTEGRAL    Likelihood = -7.17    Transmembrane   73 - 89 (  72 - 94)
   INTEGRAL    Likelihood = -6.00    Transmembrane   13 - 29 (   9 - 33)
10  INTEGRAL    Likelihood = -2.71    Transmembrane  180 - 196 ( 179 - 197)
   INTEGRAL    Likelihood = -0.59    Transmembrane  112 - 128 ( 109 - 128)

----- Final Results -----
          bacterial membrane --- Certainty=0.6265(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15         bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:AAF42284 GB:AE002544 cadmium resistance protein [Neisseria
meningitidis MC58]
20  Identities = 201/208 (96%), Positives = 205/208 (97%)

Query: 1  MRCFMIQNVVTSIIILYSGTAVDLLIIILMLFFAKRKS RKDIINIYLGQFLGSVSLILLSLL 60
          MRCFMIQNVVTSIIILYSGTAVDLLIIILMLFFAKRKS RKDIINIYLGQFLGSVSLILLSLL
25  Sbjct: 1  MRCFMIQNVVTSIIILYSGTAVDLLIIILMLFFAKRKS RKDIINIYLGQFLGSVSLILLSLL 60

Query: 61  FAFVLDYIPSKIELGLLGLIPIFLGLKVLVLLGDS DGEATAKEGLSKDNKNLIFLVAMITF 120
          FAFVLDYIPSKIELGLLGLIPI LG+KVLVLLGDS DGEATAKEGL KDNKNLIFLVAMITF
25  Sbjct: 61  FAFVLDYIPSKIELGLLGLIPIILLGLIKVLLVLLGDS DGEATAKEGLRKNKNLIFLVAMITF 120

Query: 121  ASGCADNIGVFPVYFTTLNLANLIVALLTFLVM IYLLVFS AQKLAQVPSVGETLEKYSRW 180
          ASGCADNIGVFPVYFTTLNLANLIVALLTFLVM IYLLVFS AQKLAQVPSVGETLEKYSRW
30  Sbjct: 121  ASGCADNIGVFPVYFTTLNLANLIVALLTFLVM IYLLVFS AQKLAQVPSVGETLEKYSRW 180

Query: 181  FIAVVYLGLGMYILIENNSFDMLWAVLG 208
          F+AVVVYLG LG+YIL+ENNSFDMLW VLG
35  Sbjct: 181  FVAVVYLGLGIYILVENNSFDMLWTVLG 208

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 71/200 (35%), Positives = 130/200 (64%), Gaps = 4/200 (2%)
40  Query: 1  MGQTIISAIGVYISTSIDYLVLIILFAQLSQNKQKWHIYAGQYLGTLVNGASLVAAAYV 60
          M Q ++++I +Y T++D LI+L++ FA+ K +IY GQ+LG+ L+ SL+ A+V
          Sbjct: 5  MIQNVVTSIIILYSGTAVDLLIIILMLFFAKRKS RKDIINIYLGQFLGSVSLILLSLLFAFV 64

45  Query: 61  VNFVPEAWMVGLLGLIPIYLGIRFAIVGEGEEEEEEIIEERLEQSKANQLFWTVTLTIA 120
          ++++P ++GLLGLIPI+LG++ ++G+ + E + E L + N +P V ++T A
          Sbjct: 65  LDYIPSKIELGLLGLIPIFLGLKVLVLLGDS DGEATAK--EGLSKDNKNLIF-LVAMITFA 121

50  Query: 121  S-GGDNLGIYIPYFASLDWSQTLVLLVFAIGTIIIFCELSWVLSIPLISETIEKYQRRI 179
          S G DN+G+++PYF +L+ + +V LL F + I + + L+ +P + ET+EKY R
          Sbjct: 122  SCGADNIGVFPVYFTTLNLANLIVALLTFLVM IYLLVFS AQKLAQVPSVGETLEKYSRW 181

          Query: 180  VPLVFIPLGLYIMYESGTIE 199
          + +V++ LG+YI+ E+ + +
55  Sbjct: 182  IAVVYLGLGMYILIENNSFD 201

```

SEQ ID 6150 (GBS174) was expressed in and purified from *E.coli*. The purified protein is shown in lane 7 of Figure 223.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2237-

Example 1993

A DNA sequence (GBSx2103) was identified in *S.agalactiae* <SEQ ID 6153> which encodes the amino acid sequence <SEQ ID 6154>. This protein is predicted to be Pgm. Analysis of this protein sequence reveals the following:

```

5   Possible site: 53
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.4324 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:CAB96418 GB:AJ243290 phosphoglucosyltransferase [Streptococcus thermophilus]
   Identities = 65/76 (85%), Positives = 71/76 (92%)

   Query: 1  MTYTENLQKWLDFEQLPDYLRQELLSMDEKTKEDAFYTNLEFGTAGMRGYIGAGTNRINI 60
             M+YTEN QKWLDF +LP YLR EL+SMDEKTKEDAFYTNLEFGTAGMRG IGAGTNRINI
   Sbjct: 1  MSYTENYQKWLDFAEELPAYLRDELVSMDEKTKEDAFYTNLEFGTAGMRGLIGAGTNRINI 60

20  Query: 61 YVVRQATEGLAKLIET 76
             YVVRQATEGLA+LI++
   Sbjct: 61 YVVRQATEGLAQLIDS 76

```

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6155> which encodes the amino acid sequence <SEQ ID 6156>. Analysis of this protein sequence reveals the following:

```

   Possible site: 53
   >>> Seems to have no N-terminal signal sequence

30  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4324 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

35 An alignment of the GAS and GBS proteins is shown below.

```

   Identities = 75/76 (98%), Positives = 75/76 (98%)

   Query: 1  MTYTENLQKWLDFEQLPDYLRQELLSMDEKTKEDAFYTNLEFGTAGMRGYIGAGTNRINI 60
             MTYTEN QKWLDFEQLPDYLRQELLSMDEKTKEDAFYTNLEFGTAGMRGYIGAGTNRINI
40  Sbjct: 1  MTYTENFQKWLDFEQLPDYLRQELLSMDEKTKEDAFYTNLEFGTAGMRGYIGAGTNRINI 60

   Query: 61 YVVRQATEGLAKLIET 76
             YVVRQATEGLAKLIET
45  Sbjct: 61 YVVRQATEGLAKLIET 76

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1994

50 A DNA sequence (GBSx2104) was identified in *S.agalactiae* <SEQ ID 6157> which encodes the amino acid sequence <SEQ ID 6158>. This protein is predicted to be a membrane protein. Analysis of this protein sequence reveals the following:

```

   Possible site: 53
   >>> Seems to have a cleavable N-term signal seq.
55  INTEGRAL    Likelihood = -6.21    Transmembrane    94 - 110 ( 93 - 115)
   INTEGRAL    Likelihood = -4.14    Transmembrane    172 - 188 ( 166 - 188)

```

-2238-

INTEGRAL Likelihood = -1.97 Transmembrane 130 - 146 (129 - 149)
 INTEGRAL Likelihood = -0.16 Transmembrane 62 - 78 (62 - 79)

----- Final Results -----

5 bacterial membrane --- Certainty=0.3484(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAA80247 GB:Z22520 membrane protein [Bacillus acidopullulyticus]
 Identities = 47/185 (25%), Positives = 80/185 (42%), Gaps = 23/185 (12%)

Query: 1 MKKKNKSSNIAIIAIFFAIMLVHFLSSFISSFVLVPIKPTLMHIPVIIASIAYGPRIGA 60
 MKK +I I + A+ +++ T+MHIP II I GP +G

15 Sbjct: 1 MKKSLTVRDIIVAGVLGAVAILLGVTSLGVIPVPTAAGNATIMHIPAIIIGGIMQGPVVGL 60

Query: 61 TLGALMGGISVANSSIVLLPTSYLSPFVENGNFYSLIIALVPRILIGIIPYFVYKLLHN 120
 +GA+ G S N+++ L F +++++PR+ IG++ + VY +

20 Sbjct: 61 IVGAIFGISSFLNATVPL-----FKDPLVSILPRLFIGVVAWLVIYIGIRR 105

Query: 121 R---FGLAISGAIGSLTNTVFLSGIFIFSSTYNGNIKIMLAGIISNSLAEMVIAAII 177
 + + +S IG+LTNT VL+ F + +A +N L E V+ I+

25 Sbjct: 106 KSEYVAVGLSAFIGTLTNTALVLA--MAVFRHYLTAGVAWTVVA---ITNGLPEAVVGTIV 160

Query: 178 VYLTV 182
 V

Sbjct: 161 TLAVV 165

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6159> which encodes the amino acid
 sequence <SEQ ID 6160>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have an uncleavable N-term signal seq

35 INTEGRAL Likelihood = -8.97 Transmembrane 18 - 34 (10 - 41)
 INTEGRAL Likelihood = -7.43 Transmembrane 170 - 186 (160 - 191)
 INTEGRAL Likelihood = -5.63 Transmembrane 96 - 112 (94 - 117)
 INTEGRAL Likelihood = -4.67 Transmembrane 140 - 156 (131 - 158)
 INTEGRAL Likelihood = -3.66 Transmembrane 64 - 80 (63 - 84)
 INTEGRAL Likelihood = -0.22 Transmembrane 39 - 55 (39 - 55)

40 ----- Final Results -----

bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the databases:

>GP:CAA80247 GB:Z22520 membrane protein [Bacillus acidopullulyticus]
 Identities = 47/193 (24%), Positives = 86/193 (44%), Gaps = 28/193 (14%)

50 Query: 8 RKSADISRIAIFFAIMLVHIFVSSLVFNWPIPI---KPTLVHIPVIIASVLYGPRIGAI 64
 +KS + I I + V + P+P T++HIP II ++ GP +G I

Sbjct: 2 KKSLLTVRDIIVAGVLGAVAILLGVTSLGVIPVPTAAGNATIMHIPAIIIGGIMQGPVVGLI 61

Query: 65 LGGLMGIISVITNTIILLPTNYLSPFVDHGTFFASLIITAIIPRILIGITPYCYKLIIPNQ 124
 +G + GI S + T+ L F +++I+PR+ IG+ + Y I +

55 Sbjct: 62 VGAIIFGISSFLNATVPL-----FKDPLVSILPRLFIGVVAWLVIYIGIRRK 106

Query: 125 FGLIVSGI---IGSLTNTIFVLS-GIFFFATVFDGNIKALLTAIISNAIVEMIISAIL 180
 + G+ IG+LTNT VL+ +F + T + + +N + E ++ I+

60 Sbjct: 107 SEYVAVGLSAFIGTLTNTALVLAMAVFRHYLTA-----GVAVTVAITNGLPEAVVGTIV 160

Query: 181 TFVLIPTLSRLKR 193
 T ++ ++ R

Sbjct: 161 TLAVVLAWKQIGR 173

-2239-

An alignment of the GAS and GBS proteins is shown below.

Identities = 121/184 (65%), Positives = 157/184 (84%)

```

5  Query: 6  KSSNIAIIAIFFAIMLVIIHFLSSFIFSWLVPKPTLMHIPVIIASIAYGPRIGATLGAL 65
      KS++I+ IAIFFAIMLVIIHF+SS +F+ W +PIKPTL+HIPVIIAS+ YGPRIGA LG L
      Sbjet: 9  KSADISRIAIFFAIMLVIIHFVSSLVFNWPIPIKPTLVHIPVIIASVLYGPRIGAILGGL 68

10 Query: 66  MGGISVANSSIVLLPTSYLESPFVENGNFYSLIIALVPRILIGIIPYFVYKLLHNRFGLA 125
      MG ISV ++I+LLPT+YLFSPFV++G F SLIIA++PRILIGI PY+ YKL+ N+FGL
      Sbjet: 69  MGIISVITNTIILLPTNYLFSPFVDHGTAFSLIIAIPRILIGITPYCYKLIPINQFGLI 128

15 Query: 126  ISGAIGSLTNTVFLSGGIFIFFSSTYNGNIKMLLAGIISNSLAEMVIAAIIIVYLTVPRI 185
      +SG IGSLTNT+FVLSGGIFIFF++ ++GNIK +L IISN++ EM+I+AI ++ +P +
      Sbjet: 129  VSGIIGSLTNTIFVLSGGIFIFFATVFDGNIKALLTAIISNAIVEMIISAIITFVLIPTL 188

      Query: 186  INIK 189
      +K
      Sbjet: 189  SRLK 192
  
```

20 A related GBS gene <SEQ ID 8949> and protein <SEQ ID 8950> were also identified. Analysis of this protein sequence reveals the following:

```

25 Lipop: Possible site: -1  Crend: 5
      McG: Discrim Score: 13.42
      GvH: Signal Score (-7.5): -1.93
      Possible site: 53
      >>> Seems to have a cleavable N-term signal seq.
      ALOM program count: 2 value: -6.21 threshold: 0.0
      INTEGRAL Likelihood = -6.21 Transmembrane 94 - 110 ( 93 - 115)
      INTEGRAL Likelihood = -0.16 Transmembrane 62 - 78 ( 62 - 79)
30 PERIPHERAL Likelihood = 1.70 123
      modified ALOM score: 1.74

      *** Reasoning Step: 3

35 ----- Final Results -----
      bacterial membrane --- Certainty=0.3484(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
  
```

40 The protein has homology with the following sequences in the databases:

```

      ORF01561(301 - 723 of 1017)
      EGAD|38021|39600(1 - 129 of 183) hypothetical membrane protein {Bacillus acidopullulyticus}
      GP|806536|emb|CAA80247.1||Z22520 membrane protein {Bacillus acidopullulyticus}
      %Match = 7.6
45 %Identity = 29.7 %Similarity = 53.9
      Matches = 38 Mismatches = 57 Conservative Sub.s = 31

      162      192      222      252      282      312      342      372
      KKIGYQEIEPRISLLACGDTGQALADISTILKCIQEVAN*AVNLYTISSLI*GVIMKKKNKSSNIAIIAIFFAIMLVIIH
50      |||      :| |      :: | : ::
      MKKSLTVRDIIVAGVVGAVAILLG
      10      20

      402      432      462      492      522      552      582      612
55 FLSSFIFSWLVPKPTLMHIPVIIASIAYGPRIGATLGALMGGISVANSSIVLLPTSYLESPFVENGNFYSLIIALVPR
      :||| | | | | | : | : || : | | | : : | | | : : : | | : : : | |
      VTRLGYIPVPTAAGNATIMHIPAIIIGGIMQGPVVGGLIVGAIFGIISSFLNATVPL-----FKDPLVSILPR
      40      50      60      70      80

60 642      663      693      723      753      783      813      843
      ILIGIIPYFVY---KLLHNRFGLAISGAIGSLTNTVFLVXSGGIFIFFSSTYNGNIKMLLAGIISXNSLAEMVIAAIIIVYLT
      :| : : : : | : : : | : | : | : | : | :
      LFIGVVAVLVYIGIRRKSEYVAVGLSAFIGTLTNTALVLAIVFRHYLTAGVAWTVAITNGLPEAVVGTVITVLAIVLAWK
      100      110      120      130      140      150      160
  
```

-2240-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1995

- 5 A DNA sequence (GBSx2105) was identified in *S.agalactiae* <SEQ ID 6161> which encodes the amino acid sequence <SEQ ID 6162>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence (or aa 1-18)

10 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.0165(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC44502 GB:U48885 DNA/pantothenate metabolism flavoprotein

[Streptococcus mutans]

Identities = 101/145 (69%), Positives = 122/145 (83%)

20 Query: 1 MIKRITLAVTGSISAYKAADLTSQLTKIGYDVHIIMTQAATEFITPLTLQVLSKNPIHLD 60

M K+I LAV+GSI+AYKAADL+ QLTk+GY V++ MT AA +FI PLTLQVLSKNP++ +

Sbjct: 1 MTKKILLAVSGSIAAYKAADLSHQTLKLGHYHVNVMFTNAAKQFIPPLTLQVLSKNPVYSN 60

25 Query: 61 VMDEHNPKIINHIELAKRTDLFIVAPASANTIAHLAYGFADNIVTSVALAMPDET PKLIA 120

VM E +P++INHI LAK+ DLF++ PASANT+AHLA+GFADNIVTSVALA+P E PK A

Sbjct: 61 VMKEDDPQVINHIALAKQADLFLLPPASANTLAHLAHGFADNIVTSVALALPLEVPKFFA 120

Query: 121 PAMNTKMYHNTITQRNIDILKKIGY 145

PAMNTKMY N ITQ NI +LKK GY

30 Sbjct: 121 PAMNTKMYENPITQSNITLLKKFGY 145

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6163> which encodes the amino acid sequence <SEQ ID 6164>. Analysis of this protein sequence reveals the following:

Possible site: 18

35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0076(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 146/178 (82%), Positives = 155/178 (87%)

45 Query: 1 MIKRITLAVTGSISAYKAADLTSQLTKIGYDVHIIMTQAATEFITPLTLQVLSKNPIHLD 60

M K ITLAV+GSISAYKAADLTSQLTKIGYDVHIIMTQAAT+FITPLTLQVLSKN IHL D

Sbjct: 1 MTKHITLAVSGSISAYKAADLTSQLTKIGYDVHIIMTQAATQFITPLTLQVLSKNAIHLD 60

50 Query: 61 VMDEHNPKIINHIELAKRTDLFIVAPASANTIAHLAYGFADNIVTSVALAMPDET PKLIA 120

VMDEH+PK+INHIELAKRTDLFIVAPASANTIAHLAYGFADN+VTSVALA+P TP K L I A

Sbjct: 61 VMDEHDPKVINHIELAKRTDLFIVAPASANTIAHLAYGFADNLVTSVALALPATTPK L I A 120

Query: 121 PAMNTKMYHNTITQRNIDILKKIGYQIEPRISLLACGDTGGALADISTILKCIQEV 178

PAMNTKMY N ITQ NI L IG+ EI P+ SLLACGD G GALADI IL I +

55 Sbjct: 121 PAMNTKMYQNPITQENIKRLSTIGFTEIPKSSLLACGDKGPGALADIDVILATIDTI 178

SEQ ID 6162 (GBS236) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 5; MW 21.6kDa).

-2241-

Purified GBS236-GST is shown in Figure 208 (lane 6) and in Figure 225 (lanes 4-5).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1996

- 5 A DNA sequence (GBSx2106) was identified in *S.agalactiae* <SEQ ID 6165> which encodes the amino acid sequence <SEQ ID 6166>. This protein is predicted to be pantothenate metabolism flavoprotein homolog (dfp). Analysis of this protein sequence reveals the following:

Possible site: 13
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2325(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15

A related GBS nucleic acid sequence <SEQ ID 9835> which encodes amino acid sequence <SEQ ID 9836> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AAG39941 GB:AF301375 MTW1216 [Methanothermobacter wolfeii
prophage psiM100]
Identities = 71/229 (31%), Positives = 117/229 (51%), Gaps = 27/229 (11%)

Query: 6 MKILITSGGTTEKIDTVRSITNHATGTLGKIIAEKYLREGHQVTLVTTKNAVKPESATNL 65
+++L++ GGT E ID VR ITN ++G +G +A + +G VTLV V + + L

25 Sbjct: 172 LRLVLSLGGTLEPIDFVRVITNRRSSGRMGLAVAREAYIQGADVTLVA--GTVSVDIPSQ 229

Query: 66 STFEIEDVDSLITKPLVKEHDILIHSMVSDYTPVYMADFEKVKSSDHLDTFLRKDNH 125
T E + + + L+ EHD+ + + AVSD+ PVY

30 Sbjct: 230 RTVRAETAHEMAEAVAELIGEHDVFSAAAVSDFRPVYS----- 268

Query: 126 EGKISSESEYQVLFLLKTPKVISLVKKWNPQITLVGFKLLVNVTKENLFKVARHSLIKNK 185
E KISS+SE L LK PK+I + ++ NP+ +VGFK V++E L AR + +

Sbjct: 269 EEKISSDSEI-TLRLKPNPKIIRMARETNPEAFIVGFKAEHGVSEELIAAARKQIEDSV 327

35 Query: 186 ATFILANDL-IDITSKHIIAYLLDHDNVYKATT--KEDIAQLIYEKVKK 231
A ++AND+ ++ + ++ + V + T KE++A LI ++ K

Sbjct: 328 ADMVVANDVSVEGFGSENNRAIIVSEGVTLEPTMKKEELAGLIIGEIMK 376

- 40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6167> which encodes the amino acid sequence <SEQ ID 6168>. Analysis of this protein sequence reveals the following:

Possible site: 54
>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1737(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

50 Identities = 142/230 (61%), Positives = 170/230 (73%)

Query: 4 MAMKILITSGGTTEKIDTVRSITNHATGTLGKIIAEKYLREGHQVTLVTTKNAVKPESAT 63
M MK++ITSGGTTE ID VR ITNH+TG LGK+I E++L+ H VTLVTTK A KP

55 Sbjct: 1 MTMKLITSGGTTEPIDAVRGITNHSTGQLGLITERFLQYHHDVTLVTTKTATKPLPNK 60

-2242-

Query: 64 NLSTFEIEDVDSLITLKLPLVKEHDILIHSMVSDYTPVYMADEFKVKSSDHLDTFLRKD 123
 L E+E V+ L+ LK V HDILIHSMVSDYTPVYM D E+V +D+L+ FL +
 Sbjct: 61 RLRIIEVETVNDLMAALKDQVPFHDILIHSMVSDYTPVYMTDLEQVSQADNLNCFLEH 120

5 Query: 124 NHEGKISSESEYQVLFLLKTPKVISLVKKWNPQITLVGFKLLVNVTKENLFKVARHSLIK 183
 N E KISS S+YQVLFLLKTPKVIS VK+WNP I LVGFKLLVNV +E L KVAR SL K
 Sbjct: 121 NSEPKISSASDYQVLFLLKTPKVISYVKQWNPNIKLVGFKLLVNVVQEELIKVARASLAK 180

10 Query: 184 NKATFILANDLIDITSKHHIAYLLDHDNVYKATTKEDIAQLIYEKVKKYD 233
 N A +ILANDL+DI + H A L+ ++ V A TKE IA L+YE++ K+D
 Sbjct: 181 NHADYILANDLVDIQTGMHKALLISNNEVASADTKEAIADLLYERMTKHD 230

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 1997

A DNA sequence (GBSx2107) was identified in *S.agalactiae* <SEQ ID 6169> which encodes the amino acid sequence <SEQ ID 6170>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence
 20 INTEGRAL Likelihood = -0.22 Transmembrane 117 - 133 (117 - 133)

----- Final Results -----
 bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9833> which encodes amino acid sequence <SEQ ID 9834> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:BAB07541 GB:AP001520 unknown conserved protein in B. subtilis
 [Bacillus halodurans]
 Identities = 94/221 (42%), Positives = 133/221 (59%), Gaps = 2/221 (0%)

35 Query: 52 AEKPFITWTEVFLREINRSNQEIIHLIWPMTKTIVILGMLDRELPHLELAKKEIISRGYEPV 111
 A + F + + I +S L W TV+LG+ D LP ++ + + ++ +
 Sbjct: 27 ALQSFAYDDTLCTSIGKSQSPPTLRAWVHNTTVLGIQDSRLPQIKAGIEALKGFQHDVI 86

Query: 112 VRNFGGLAVVADEGILNFSVLVDPVFERKLSISDGYLIMVDFIRSIFSDFYQPIEHFEVE 171
 VRN GGLAVV D GILN SLV+ + E+ SI DGY +M + I S+F D + IE E+
 40 Sbjct: 87 VRNSGGLAVVLDGILNLSLVLKE--EKGFSIDGYELMYELICSMFQDHREQIEAREIV 144

Query: 172 TSYCPGKFDLSINGKKFAGLAQRRIKNGIAVSIYLSVCGDQKGRSQMISDFYKIGLGDTG 231
 SYCPG +DLSI+GKKFAG++QRR I+ G+AV IYL V G R++MI FY +
 45 Sbjct: 145 GSYCPGSYDLSIDGKKFAGISQRRIRGGVAVQIYLCVSGSGAERAKMIRTFYDKAVAGQP 204

Query: 232 SPIAYPNVDPEIMANLSDLLDCPMTVEDVIDRMLISLQVG 272
 + YP + PE MA+LS+LL P V DV+ + L++L+Q G
 Sbjct: 205 TKFVYPRIKPETMASLSSELLGQPHNVSDVLLKALMTLQQHG 245

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6171> which encodes the amino acid sequence <SEQ ID 6172>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have an uncleavable N-term signal seq
 55 INTEGRAL Likelihood = -0.22 Transmembrane 95 - 111 (95 - 111)

----- Final Results -----
 bacterial membrane --- Certainty=0.1086(Affirmative) < succ>

-2243-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP:BA07541 GB:AP001520 unknown conserved protein in B. subtilis
 [Bacillus halodurans]
 Identities = 97/228 (42%), Positives = 138/228 (59%), Gaps = 2/228 (0%)

10 Query: 30 ALSPFVWTEVFLKTIHQEPNQLILHIWPMTRTVILGMLDRQLPYFELAKTEIGNNGYVPV 89
 AL F + + +I + + L W TV+LG+ D +LP + + + +
 Sbjct: 27 ALQSFAYDDTLCTSIGKSQSPPTLRVWVHNTTVLGIQDSRLPQIKAGIEALKGFQHDVI 86

15 Query: 90 TRNIGGLAVVADDGILNFSVLIPDHFSEISISNAYLIMVDVIRESFSDYYQRIEYHEIK 149
 RN GGLAVV D GILN SLV+ + + SI + Y +M ++I F D+ ++IE EI
 Sbjct: 87 VRNSGGLAVVLDGILNLSVLKEE--KGFSIDGYELMYELICSMFQDHREQIEAREIV 144

20 Query: 150 NSYCPGNFDLSIAGRKFAGIAQRRIKGIIVSVYLSVCGDQAARGQLIKDFYEAGTQGEV 209
 SYCPG++DL SI G+KFAGI+QRRI+ G+ V IYL V G A R ++I+ FY+ G+
 Sbjct: 145 GSYCPGSYDLSIDGKKFAGISQRRIRGGVAVQIYLCVSGSGAERAKMIRTFYDKAVAGQP 204

Query: 210 TKVNPQIDPECMATLSELLETPFTVAEVLRLRLTLRQLGFSLTEKS 257
 TK YP+I PE MA+LSELL P V++VL + +TL+Q G SL +S
 Sbjct: 205 TKFVYPRIKPETMASLSELLGQPHNVSDVLLKALMTLQHGASLLTES 252

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 155/275 (56%), Positives = 199/275 (72%), Gaps = 8/275 (2%)

30 Query: 32 QDLAQLPVSIKDYVTDAQDAEKPFITWTEVFLREINRSNQEILHIWPMTKTVILGMLDR 91
 +DLA LP+ ++ D A PF+WTEVFL+ IN+ ++ILHIWPMT+TVILGMLDR
 Sbjct: 10 RDLASLPFVYGDGNKKVPGALSPFVWTEVFLKTIHQEPNQLILHIWPMTRTVILGMLDR 69

35 Query: 92 ELPHLELAKKEIISRGYEPVVRNFGGLAVVADEGILNFSVLIPDVFERKLSISDGYLIMV 151
 +LP+ ELAK EI + GY PV RN GGLAVVAD+GILNFSVLIPD F +SIS+ YLIMV
 Sbjct: 70 QLPYPFELAKTEIGNNGYVPVTRNIGGLAVVADDGILNFSVLIPDHFSEISISNAYLIMV 129

40 Query: 152 DFIRSIFSDFYQPIEHFEVETSYCPGKFDLSINGKKFAGLAQRRIKNGIAVSIYLSVCGD 211
 D IR FSD+YQ IE+ E++ SYCPG FDL SI G+KFAG+AQRIK GI VSIYLSVCGD
 Sbjct: 130 DVIRESFSDYYQRIEYHEIKNSYCPGNFDLSIAGRKFAGIAQRRIKGIIVSVYLSVCGD 189

45 Query: 212 QKGRSQMISDFYKIGLGTGSPAIYPNVDP EIMANLSDLLDCPMTVEDVIDRMLISLKQV 271
 Q R Q+I DFY+ G + + YP +DPE MA LS+LL+ P TV +V++R+ ++L+Q+
 Sbjct: 190 QAARGQLIKDFYEAGTQGEVTKVNPQIDPECMATLSELLETPFTVAEVLRLRLTLRQL 249

Query: 272 GFN-----DRLLMIRPDLVAEFNRFQAKSMANKG 300
 GF+ D+ L+ D V + R Q + + +G
 Sbjct: 250 GFSLTEKSPDQALLTNFDAV--YERMQLLEVVRKEG 282

A related GBS gene <SEQ ID 8951> and protein <SEQ ID 8952> were also identified. Analysis of this protein sequence reveals the following:

50 Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: -16.85
 GvH: Signal Score (-7.5): -5.07
 Possible site: 49

55 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 1 value: -0.22 threshold: 0.0
 INTEGRAL Likelihood = -0.22 Transmembrane 117 - 133 (117 - 133)
 PERIPHERAL Likelihood = 0.47 73
 modified ALOM score: 0.54

60 *** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2244-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

```

ORF01564(451 - 1116 of 1518)
EGAD|13388|BS3758(27 - 249 of 281) hypothetical 31.4 kd protein in pta 3'region {Bacillus subtilis} OMNI|NT01BS4391 hypothetical protein SP|P39648|YWFL_BACSU HYPOTHETICAL 31.4 KDA PROTEIN IN PTA 3'REGION. GP|414014|emb|CAA51646.1||X73124 ipa-90d {Bacillus subtilis} GP|2636300|emb|CAB15791.1||Z99123 alternate gene name: ipa-90d {Bacillus subtilis} PIR|S39745|S39745 ywFL protein - Bacillus subtilis
%Match = 15.8
%Identity = 40.8 %Similarity = 61.0
Matches = 91 Mismatches = 82 Conservative Sub.s = 45

321          351          381          411          441          471          501          531
*WNLRETYWKTSSDCDKINLAESRERMSDLLLEWQDLAQLPVSIFKDYVTTAQDAEKPFTWTEVFLREINRSNQEIIILHI
||::| : : :
MANQPIDLLMQPKWRVIDQSSSLGPLFDAKQSFAMDDTLCMSVGVKGVSPATARS
10 20 30 40 50

561          591          621          651          681          711          738          768
WPMTKTIVILGMLDRELPHLELAKKEIISRGYEPVVRNFGGLAVVADEGILNFSLVIPDVFERK--LSISDGYLIMVDIFRS
|::||:| || : : ||:| ||||| |:|:| ||:| |:|:| || ||:::|
VVHHDITIVLGIQDTRLPLFLQDGISLLESEGYRVIVRNSGGLAVVLDGVLNISLIFED--EKKGIDIDKGYEAMVELMRR
70 80 90 100 110 120 130

798          828          858          888          918          972          996
IFSDFYQPIEHFEVETSYCPGKFDLSINGKKFAGLAQRRIKNGIAVSIYLSVCGDQKG--RSQMISDFYKIGLGD--TGS
:: : ||:|:| ||||| :|||||||:::|::: |:| ||| |:| |:|:| ||:| |
MLRPYNAKIEAYEIEGSYCPGSYDLSINGKKFAGISQRRVRGGVAVQIYL--CADKSGSERADLIRRFYQAALKDKQNDK
150 160 170 180 190 200

1026          1056          1086          1116          1146          1176          1206          1236
PIAYPNVDPEIMANLSDLLDCPMTVEDVIDRMLISLKQVGFNDRLLMIRPDLVAEFNRFQAKSMANKGMVSRDE*CPR*F
||:| |||:|:| :||:|:| :| ||:
KGVYPEIRPTMASLSELLQKDISVQDLMFALLTELKALSTHLYSAGLSIDEEMFEKKNLVRMAERNAKVFG
220 230 240 250 260 270 280

```

SEQ ID 8952 (GBS390) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 7; MW 37kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 82 (lane 3; MW 62kDa).

40 GBS390-GST was purified as shown in Figure 216, lane 12.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1998

45 A DNA sequence (GBSx2108) was identified in *S.agalactiae* <SEQ ID 6173> which encodes the amino acid sequence <SEQ ID 6174>. This protein is predicted to be probable trimethylamine dehydrogenase (nemA). Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
```

```
50      ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2218(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

55 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA83700 GB:Z33015 similar to trimethylamine DH [Mycoplasma

-2245-

capricolum]

Identities = 162/311 (52%), Positives = 219/311 (70%), Gaps = 1/311 (0%)

5 Query: 3 NVQGNLFRPLTLPLNGLSLENRFVLSPMVNSSTSEGFVTDDDIAYAVRRAKSAPLQITGA 62
N LF P L NG LENRFVLSPM + +T +G +TD + Y RR+ SAPLQITG
Sbjct: 2 NKYEKLFEFPYL-NGFKLENRFVLSPMTLSLATLDGKITDKEADYVKRRSHSAPLQITGG 60

10 Query: 63 AYITEYQQLFEYGFVSVDKEDIPGLTKLAKAMKSKGAKAVLQLTHAGRFSSHTLARHGYV 122
Y E+GQLFEYG S D+DIP LT+L + MK+ +LQL HAG+FS +L ++GY+
Sbjct: 61 VYFDEFGQLFEYGISAKSDDIPSLTRLRYQEMKTDNSNCVILQLAHAGKFSKTSCLKYGYL 120

15 Query: 123 YGPSPMQLQSPYPHQVKELTHKDILRIIDEYVQATRRAIQAGFDGVEISSAQRLLIQTFF 182
YGPS + +P H+V EL + I +II +Y AT R I+AGF+G+EIS AQRLLIQTFF
Sbjct: 121 YGPSYEKNHTPIEHEVLELPKEKIKQIIQDYKDATLRVIKAGFNGIETISMAQRLLIQTFF 180

20 Query: 183 STFSNQKDEYGPQTLTNRCRLGLEVFKAQKQVIREEAESDFILGFRATPEETRGSQIGY 242
S N+R DEY NR R LEV KA+++VI + A +FI GFRATPEET G +GY
Sbjct: 181 SQIINKRTDEYSATNFENRSRRCLEVVKAIREVIDKYAPKNFIFGFRATPEETYGDILGY 240

25 Query: 243 SIEEFMEFLEKILAIQAQVDYLAIASWGHDFVRNTIRSEGVYKQQLVNQVIFEHFGDRVPI 302
+IE+F++ ++KI+ I ++ YLAIASWGH++ N +RS YKQQLVN+VI++ + +++PI
Sbjct: 241 TIEDFIQLVDKIIIEIGKISYLAIASWGHDIYLNKVRNTPKYKQQLVNQVIYDIYKNKLPI 300

Query: 303 MATGGINSASK 313
+++GGIN+ +K
Sbjct: 301 ISSGGINTPTK 311

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6175> which encodes the amino acid sequence <SEQ ID 6176>. Analysis of this protein sequence reveals the following:

30 Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
35 bacterial cytoplasm --- Certainty=0.3055(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 265/390 (67%), Positives = 321/390 (81%)

40 Query: 8 LFRPLTLPLNGLSLENRFVLSPMVNSSTSEGFVTDDDIAYAVRRAKSAPLQITGAAYITE 67
LF PLTLPLNG L+NRFLVLSPMVNSST +G+VT DD++YA+RRA SAPLQITGAAY+
Sbjct: 8 LFEPLTLPLNGSQLDNRFVLSPMVNSSTKDGYVTQDDVSYALRRAASAPLQITGAAYVDP 67

45 Query: 68 YGQLFEYGFVSVDKEDIPGLTKLAKAMKSKGAKAVLQLTHAGRFSSHTLARHGYVYGPSP 127
YGQLFEYGFVSVDK DI GL +LA+AMK+KGAKAVLQLTHAGRF+SH L ++G+VYGPS
Sbjct: 68 YGQLFEYGFVSVDKADISGLKELAQAMKAKGAKAVLQLTHAGRFASHALTKYGFVYGPSY 127

50 Query: 128 MQLQSPYPHQVKELTHKDILRIIDEYVQATRRAIQAGFDGVEISSAQRLLIQTFFSTFSN 187
MQL+SP PH+VK LT + I +I Y QATRRAIQAGFDGVE+SSAQRLLIQTFFSTFSN
Sbjct: 128 MQLRSPQPEHVKPLTGQQIEELIAAYAQAATRRAIQAGFDGVEVSSAQRLLIQTFFSTFSN 187

55 Query: 188 QRKDEYGPQTLTNRCRLGLEVFKAQKQVIREEAESDFILGFRATPEETRGSQIGYSIEEF 247
+R D YG QTL NR +L L V +AVQ+VI++EA FI GFRATPEETR+ IGYSI+EF
Sbjct: 188 KRTDSYGCQTLFNRSKLTFLAVLQAVQVVIKQEBAPDGFIFGFRATPEETRGNIDIGYSIDEF 247

60 Query: 248 MEFLEKILAIQAQVDYLAIASWGHDFVRNTIRSEGVYKQQLVNQVIFEHFGDRVPIMATGG 307
++ ++ +L +A++DYLAIASW VFRNT+RS G Y G+ VNQV+ ++ +++P+MATGG
Sbjct: 248 LQLMDWVLNVAKLDYLAIASWGRHVFRNTVRSPGPYPYGRRVNVVRDYLNRNKLPMATGG 307

Query: 308 INSASKVFEALQHAHMIGASTPLVVDPEFLQKIKAKCSDQINLRIKVSLEGLAIPKASF 367
+N+ K EAL HA IG STP VVDPEF KIK C + I+LRI+ +DL+ LAIP+ASF
Sbjct: 308 MNTPDKAIEALAHADFIGVSTPFVVDPEFAHKIKEGCEESIHLRIRPADLKS LAIPQASF 367

65 Query: 368 KDIVPLMDYGESLPKEAREVFRELRSNYRE 397

-2246-

KDIVPLMDYGESLPKE+R +FR L NY+E
 Sbjct: 368 KDIVPLMDYGESLPKESRTLFRSLTHNYKE 397

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1999

A DNA sequence (GBSx2109) was identified in *S.agalactiae* <SEQ ID 6177> which encodes the amino acid sequence <SEQ ID 6178>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3748 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04594 GB:AP001510 unknown conserved protein [Bacillus halodurans]
 Identities = 121/333 (36%), Positives = 192/333 (57%), Gaps = 12/333 (3%)

Query: 1 MKLSVLDYGLIDYGKTASDAIQETILLSSQEAERLGYHQFWVAEHHGVKAFSISNPPELMIM 60
 MKLSVLD I YG A +A+++T L++ E LGYH+FWV+EHH + S+PE++I
 Sbjct: 1 MKLSVLDQSPFIAYGSNAKEALRQTTELAKVTEALGYHRFWVSEHHDASTLAGSSPEVLIA 60

Query: 61 HLANQTKSIKIGSGGIMPLHYSSFKLAETLKTLETCHPNRVSIGLGNLGTVKVSNALRS 120
 HLA TK I++GSGG+M HYS++K+AE K LE HP R+ +GLG + G + ++
 Sbjct: 61 HLAHTKKIRLGSGGVMLPHYSAYKVAENFKLLEALHPGRIDVGLGRAPCGMPIAKMALQ 120

Query: 121 LHK---AHDYEEVLEELKSWLIDESSSKEPL---VQPTLSSFPDLYVLGSGQKSAYLAA 173
 K H Y ++++ +L D+ + P + + PD+++LGS SA +AA
 Sbjct: 121 EGKEQNIHKYPLQVKDVIGYLQDDLPTDHRFHGLKATPLIDTVPDVLGSSGGSANVAA 180

Query: 174 KLGLGFTFGVFPFMDKDLTEAKKLSSLYYHQFEYYPNKSPNLMVAAFFVVIADTSEAE 233
 + G GF F F++ + +A + Y F+ P VA FV+ ADT E+A+
 Sbjct: 181 ENGTGFAFA--HFINGEGGVQAVE---SYRETFQPSALFDRPQTSVAIFVICADTDEQAD 235

Query: 234 NIAKTLDIWMNLGNKDFNEFATFPTIEEANYQLTPEQKAKIKSNRDRMIVGDPKQVKESL 293
 IA +LD+ ++ ++ P+IE A Y +P ++A+I+ NR RMIVG PK V++ L
 Sbjct: 236 QIASSLDLSLIMLENGQLSKGTPSIESALSYPSPFERARIRENRKRMIVGSPKAVRQQL 295

Query: 294 DALVNASQAEELLILPLVPGLDQRIKSLKLLSQ 326
 L A + EE++++ + + RI+S +LL +
 Sbjct: 296 VELARAYETEEVIVVTITHRFEDIRIRSYELLGE 328

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6179> which encodes the amino acid sequence <SEQ ID 6180>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.60 Transmembrane 212 - 228 (210 - 229)

----- Final Results -----
 bacterial membrane --- Certainty=0.2041 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 173/329 (52%), Positives = 241/329 (72%), Gaps = 1/329 (0%)

Query: 1 MKLSVLDYGLIDYGKTASDAIQETILLSSQEAERLGYHQFWVAEHHGVKAFSISNPPELMIM 60

-2247-

MK+S+LDYG+ID KT +A+ ET L+Q A++LG+H+FWVAEHH + AF+IS+PEL++M
 Sbjct: 1 MKVSILDYGVIDKEKTPQEALLETRCLAQVADKLGFRFWVAEHHNIYAFAISSPELLMM 60

 Query: 61 HLANQTKSIKIGSGGIMPLHYSSFKLAETLKTLETCHPNRVSIGLGNLGTVKVSNALRS 120
 HLA+ TK I+IGSGGIMPLHYSSFK+AE + TLE HPNR+ +G+GNSLGT V AL S
 Sbjct: 61 HLADHTKQIRIGSGGIMPLHYSSFKIAEWIMTLEALHPNRIDLGTGNLGTTLVQRALSS 120

 Query: 121 LHKADYEEVLEELKSWLIDESSSKEPL-VQPTLSSFPDLYVLGSGQKSAYLAACLGLGF 179
 +H Y +V+ EL +L + S P+ V P +++P ++ L + ++A LA +LGLG+
 Sbjct: 121 IHCKDSYSQVVTELYQYLNPDHLSPLPIFVNPRGNTYPQIWTLSNSLETAELAGQLGLGY 180

 Query: 180 TFGVFPFMDKDLTEAKKLSSLYHQFPREYYPNKSPNLMVAAFVVIADTSEEAEINAKTL 239
 TFG+FP++ KDP+TEAK++S+ Y F K P L++A F+V++DT E+AE +AK L
 Sbjct: 181 TFGIFPYIPKDPITEAKRVSAHYRKAFRPSKLLKIPKLILAVFIVLSDTDEKAEALAKPL 240

 Query: 240 DIWMLGNKDFNEFATPPTIEEANYHQLTPEQKAKIKSNRDRMIVGDPKQVKESLDALVNA 299
 DIWMLG +DFNEF T+P +EEA +Y LT +Q+ I +NR RM++G P VK+ LD L+ A
 Sbjct: 241 DIWMLGQQDFNEFKTYPDVEEARNYHLTEKQREAIANRMRMIGSPHTVKKQLDRLEA 300

 Query: 300 SQAEELLPLVPGLDQRIKSLKLLSPLY 328
 QA+ELL IPLVP R ++L+LL+ LY
 Sbjct: 301 CQADELLAIPLVPEFANRQRTLELLADLY 329

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2000

A DNA sequence (GBSx2110) was identified in *S.agalactiae* <SEQ ID 6181> which encodes the amino acid sequence <SEQ ID 6182>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2384 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF81345 GB:AC007767 Identical to a glycine cleavage system
 H-protein precursor from Arabidopsis thaliana gb|P25855.
 It contains a glycine cleavage H-protein domain
 PF|01597. ESTs gb|R90208, gb|AI
 Identities = 30/91 (32%), Positives = 53/91 (57%), Gaps = 1/91 (1%)

 Query: 18 TISLTPELQDDLGTGVGYVEFTD-DANLEVDDVILNIEASKTVMAILSPITGKVVKNVNTAA 76
 TI +T QD LG V +VE + +++++ + +E+ K ILSP++G+V++VNT
 Sbjct: 59 TIGITDHAQDHLGEVVFVELPEANSSVSKEKSFGAVESVKATSEILSPISGEVIEVNTKL 118

 Query: 77 SQEPTLLNSEKADENWLVLTEVDYAAFEAL 107
 ++ P L+NS ++ W++ + A EAL
 Sbjct: 119 TESPGLINSSPYEDGWMIKVKPSSFALEAL 149

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6183> which encodes the amino acid sequence <SEQ ID 6184>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3544 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-2248-

An alignment of the GAS and GBS proteins is shown below.

Identities = 80/110 (72%), Positives = 98/110 (88%)

```

5  Query: 1  MKKIANYLLIEKNEELYTISLTPELQDDLGTVGYVEFTDDANLEVDDVILNIEASKTVMA 60
      MKKIANYLLIEK ++ YTIS+TPELQDD+GT+GY EFTD+ +L VDD+ILN+EASKTVM+
      Sbjct: 1  MKKIANYLLIEKTDDRYTISMTPELQDDIGTIGYAEFTDNDHLAVDDIILNLEASKTVMS 60

10  Query: 61  ILSPLTGKVVKVNTAASQEP TLLNSEKADENWLVLTEVDYAAFEALENA 110
      +LSPL G VV+ N AA+ P TLLNSEKA+ENW+VVL T+VD AAF+ALE+A
      Sbjct: 61  VLSPLAGAVVERNEAATLTPTLLNSEKAENWIVVLTVDVQAAFDALEDA 110

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 2001

A DNA sequence (GBSx2111) was identified in *S.agalactiae* <SEQ ID 6185> which encodes the amino acid sequence <SEQ ID 6186>. This protein is predicted to be LRP16 (b1045). Analysis of this protein sequence reveals the following:

```

20  Possible site: 17
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.0608(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
25           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:AAF15294 GB:AF202922 LRP16 [Homo sapiens]
    Identities = 73/171 (42%), Positives = 98/171 (56%), Gaps = 13/171 (7%)

30  Query: 88  DICLLQVDAIVNAANSKLLGCFIPNHHCIDNQIHTFAGSRLRLACHQLMTQQGRMEAVGQ 147
      DI L+VDAIVNAANS LLG +D IH AG L C L + + G+
      Sbjct: 78  DITKLEVDIAIVNAANSLLG-----GGVDGCIHRAAGPLLTDECR TLQ SCK-----TGK 127

35  Query: 148 AKLTESYHLFCKYVIHTVGPYVKVDQKPSRIREDLLKSSYKSC LQAVRANLKTIVFPCI 207
      AK+T Y LP KYVIHTVGP + S+ E L+S Y S L L + L+++ FPCI
      Sbjct: 128 AKITGGYRLPAKYVIHTVGP IAYGEPSASQAEE--LRSCYLSL D L L L E H R L R S V A F P C I 185

40  Query: 208 STGEFGFPNQRAAELAVQAILEWQRENQHKL-YIIFNTFTPKDQDIYQKLL 257
      STG FG+P + AA E+ + + EW +++ K+ +I F KD+DIY+ L
      Sbjct: 186 STGVFGYPCRAAAEIVLATLREWLEQHKDKVDRLLIICVFLEKDEDIYRSRL 236

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6187> which encodes the amino acid sequence <SEQ ID 6188>. Analysis of this protein sequence reveals the following:

```

45  Possible site: 25
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.1992(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
50           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 139/266 (52%), Positives = 178/266 (66%), Gaps = 6/266 (2%)

```

55  Query: 1  MPNQKQLLLAMIEY LQSEK L TDVDDL-----RTTDLQTVWRGLV NQ QDPQNISQ EYLSLED 56
      MP+ LL MI LQ+E+LT T Q +WR L+NQ+ +S++YL+LED

```

-2249-

Sbjct: 1 MPSSFLLGEMIGLLQTEQLTSSWACPLPNALTKRQDLWRALINQRPALPLSKDYLNLED 60

Query: 57 RYLSHWNTQKVKTIDVCHQTVYSNVFTYHGDICLLQVDAIVNAANSKLLGCFIPNHHCI 116
YL W + ++ C +T Y+++F YHGD L VDAIVNAANS+LLGCF PNH CI

5 Sbjct: 61 AYLDWRASFVPVSVKDCQKNTYNTSLFLYHGDIRYLAVDAIVNAANSELLGCFSPNHGCI 120

Query: 117 DNQIHTFAGSRLRLACHQLMTQQGRMEAVGQAKLTESYHLPCKYVIHTVGPYVKVDQKPS 176
DN IHTFAGSRLRLAC +MT+QGR EA+QAKLT +YHLP Y+IHTVGP + S

10 Sbjct: 121 DNAIHTFAGSRLRLACQAIMTEQGRKEAIGQAKLTSAHYLHPASYIHTVGPRIKGGHVS 180

Query: 177 RIREDLLKSSYKSCQLAVRANLKTIVFPCISTGEFGFPNORAAELAVQAILEWQRENQH 236
IR DLL Y+S L LAV+A L ++ F ISTGEFGFP + AA++A++ +L+WQ E+

Sbjct: 181 PIRADLLARCYSRLDLAVKAGLTSIAFCSISTGEFGFPKKEAAQIAIKTVLKWQAEHPE 240

15 Query: 237 K--LYIIFNTFTPKDQDIYQKLLLE 260
L IFNTFT +D+ +Y L KE

Sbjct: 241 SKTLTITFTFTSEDKALYDTYLQKE 266

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2002

A DNA sequence (GBSx2112) was identified in *S.agalactiae* <SEQ ID 6189> which encodes the amino acid sequence <SEQ ID 6190>. Analysis of this protein sequence reveals the following:

Possible site: 41
25 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2171(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6191> which encodes the amino acid sequence <SEQ ID 6192>. Analysis of this protein sequence reveals the following:

Possible site: 41
35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2477(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 218/284 (76%), Positives = 250/284 (87%)

45 Query: 4 WKTLEKTNHSQSEILSQLIEESDAIVVGIGAGMSAADGFTYIGPRFEEAFPDFTAKYQLL 63
W T + N +Q+E L+QLI+E+DA+VVGIGAGMSAADGFTYIG RFE AFPDFTAKYQ L

Sbjct: 4 WTTYPQKNLTQAEQLAQLIKEADALVVGIGAGMSAADGFTYIGSRFETAFPDFTAKYQFL 63

50 Query: 64 DMLQASLYDFEDWEEYWAFQSRFVALNYLDQPVGQAYLDLKDILAKKEYHIITNADNAF 123
DMLQASL+DFEDW+EYWAFQSRFVALNYLDQPVGQ+YLDLK+IL K+YHIITNADNAF

Sbjct: 64 DMLQASLDFEDWQBYWAFQSRFVALNYLDQPVGQSYLDLKEILGTDYHIITNADNAF 123

55 Query: 124 AVADYNLEKVFHIQGEYGLWQCSQHCHQQTyrNDQAIRQMIAQQKDMKIPSNLIPKCPKC 183
VA Y+ +FHIQGEYGLWQCSQHCHQQT++D IRQMIA+QK+MK+P LIP CP+C

Sbjct: 124 WVAGYDPHNIFHIQGEYGLWQCSQHCHQQTyKDDTVIRQMIAEQNMKVPGQLIPHCPEC 183

Query: 184 DQPFPEINKRNEEKGMEVADDFHAQRQRYENFLSQHQNDKVLYLEIGVGHTTPQFIKHPFW 243
+ PFEINKRNEEKGMEVADDFHAQ+ RYE FLS+H+ KVLYLEIGVGHTTPQFIKHPFW

-2250-

Sbjct: 184 EAPFEINKRNEEKGMVEDADFAQKARYEAFLEHKEGKVLYLEIGVGHTTPQFIKHPFW 243

Query: 244 RFVSLNENSLFVTLNHHKHYRIPQKIRSRVQLTQHIAELIAEAK 287
+ VS N N+LFVTLNHHKHYRIP IR +S++LT+HIA+LI+ K

5 Sbjct: 244 KRVSENPALFVTLNHHKHYRIPLSIRRSLELTHEIAQLISATK 287

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2003

- 10 A DNA sequence (GBSx2113) was identified in *S.agalactiae* <SEQ ID 6193> which encodes the amino acid sequence <SEQ ID 6194>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1086(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12865 GB:Z99109 similar to lipoate-protein ligase [Bacillus subtilis]
Identities = 130/331 (39%), Positives = 206/331 (61%), Gaps = 5/331 (1%)

- 25 Query: 9 NGKRITDGAIALAMQVYILQNVFLDDIILFPYYCDPKVEIGKFNQNAVIETNQEYLKEHDI 68
++ I D I LA++ Y ++++ + L Y P + IGK QN + E N +Y++E+ I
Sbjct: 5 DNQNINDPRINLAIEEYCVKHLDPQQYLLFYVNQPSIIIGKNQNTIEEINTKYVEENGI 64

- 30 Query: 69 PVVRDRTGGGAVYVDGSAVNICYLMKDHGQ-FGDFKRAYEPAIKALKTLGASSVEMRERN 127
VVRR +GGGAVY D G +N ++ KD G F +FK+ EP I+AL LG + E+ RN
Sbjct: 65 IVVRLSGGGAVYHDLGNLNFSGFITKDDGDSFHNFKKFTPEPVIQALHQLGVEA-ELSGRN 123

- 35 Query: 128 DLVIDGKKVSGAAMTIIVNGRIYGGYSLLLDVFDDAMEKVLNPNRKKIESKGIKSVRSRVG 187
D+V+DG+K+SG A GRI+ +L+ D D + L + KIESKGIKS+RSRV
Sbjct: 124 DIVVDGRKISGNAQFATKGRIFSHGTLMPDSADHVVSALKVKKDKIESKGIKSIRSRVA 183

- 40 Query: 188 DIRSHLSEDYRHITTDQFKDLMVCQLLHIDHIDQAKRYHLTEKDWAAIDALADEKYKNWD 247
+I L + +TT++F+ ++ + + + + Y LTEKDW I ++ E+Y+NWD
Sbjct: 184 NISEFLDDK---MTTEEFRSHLLRHIFNTNDVGNVPEYKLTTEKDWETIHQISKERYQNWD 240

- 45 Query: 248 WNYGNPQYSYHRDARFPGTYDFHLEIEKGIITNCRIYGDFFSSKDISDIENLLIGCPM 307
WNYG SP+++ + R+P G+ D HLE++KG I +C+I+GDFF D+S+IENLL+G
Sbjct: 241 WNYGRSPKFNLNHRSKRYPVGSIDLHLEVKKGKIEDCKIFGDFFGVGDVSEIENLLVGKQY 300

- Query: 308 KEELVLEKLSTLSLEDYFGQTSPEEIKAVLF 338
+ ++ + L ++L+ YFG + E+ +++
Sbjct: 301 ERSVIADVLEGVNLKHYFGNITKEDFLDLIY 331

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6195> which encodes the amino acid sequence <SEQ ID 6196>. Analysis of this protein sequence reveals the following:

- 50 Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 55 bacterial cytoplasm --- Certainty=0.0939(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 248/339 (73%), Positives = 283/339 (83%)

-2251-

Query: 1 MYLIEPIRNGKRITDGAIALAMQVYILQNVFLDDDLFFPYCDPKVEIGKFCQNAVETNQ 60
 MYLIEPIRNGKRITDGA+ALAMQVY+ +N+FLDDDLFFPYCDPKVEIGKFCQNAV+ETNQ
 Sbjct: 1 MYLIEPIRNGKRITDCAVALAMQVYVQENLFLDDDLFFPYCDPKVEIGKFCQNAVETNQ 60

Query: 61 EYLKEHDIPVVRDRTGGGAVYVDSGAVNICYLMKDHGQFGDFKRAYEPAIKALKTLGASS 120
 EYLKEH IPVVRDRTGGGAVYVDSGAVNICYL+ D+G FGDFKR Y+PAI+AL LGA+
 Sbjct: 61 EYLKEHHIPVVRDRTGGGAVYVDSGAVNICYLINDNGIFGDFKRQYPAIEALHHLGATE 120

Query: 121 VEMRERNDLVIDGKKVSGAAMTI VNGRIYGGYSLLLDVFDFAMEKVLNPNRKKIESKGIK 180
 VEM RNDLVIDGKKVSGAAMTI NGR+YGGYSLLLDVFDF+AMEK L PNRKKIESKGI+
 Sbjct: 121 VEMSGRNDLVIDGKKVSGAAMTIANGRVYGGYSLLLDVFDFEAMEKALKPNRKKIESKGIK 180

Query: 181 SVRSRVGDIRSHLSEDIRHITTDQFKDLMVCQLLHIDHIDQAKRYHLTEKWAAIDALAD 240
 SVRSRVG+IR HL+ Y+ IT ++FKDLMVCQLL I+ I QAKRY LTEKDW IDAL +
 Sbjct: 181 SVRSRVGNIREHLAPQYQGITIEEFKDLMLVCQLLQIETISQAKRYDLTEKDWQIDALTE 240

Query: 241 EKYKNWDWNYGNPQYSYHRDARPPSGTYDFHLEIEKGIITNCRIYGDFFSSKDISDIEN 300
 KY NW+WNYGN+PQY YHRD RF GT D HL+I+KG I CRIYGDFF DI+++E
 Sbjct: 241 RKYHNWENWNYGNAPQYRYHRDGRFTGGTVDIHLDIKKGYIAACRIYGDFFGKADIAELEG 300

Query: 301 LLIGCPMKEELVLEKLSTLSLEDYFGQTSPEEIKAVLFS 339
 LIG M++E VL L+ + L Y G + EE+ ++FS
 Sbjct: 301 HLIGTRMEKEDVLATLNAIDLAPYLGAITAEELGDLIFS 339

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2004

A DNA sequence (GBSx2114) was identified in *S.agalactiae* <SEQ ID 6197> which encodes the amino acid sequence <SEQ ID 6198>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.49 Transmembrane 196 - 212 (196 - 212)

----- Final Results -----
 bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB49329 GB:U39612 formyl-tetrahydrofolate synthetase
 [Streptococcus mutans]
 Identities = 432/556 (77%), Positives = 493/556 (87%)

Query: 1 MKTDIEIAQSVALKPIAEIVEQVGIGFDDIELYGKYAKLSFDKIEAVKSQKVGLILVT 60
 MKTDIEIAQSV L+PI +V+++GI FDD+ELYGKYAKL+FDKI+AV+ GKL+LVT
 Sbjct: 1 MKTDIEIAQSVDLRPTNVVKKLGIDFDDLELYGKYAKLTFDKIKAVEENAPGKLVLT 60

Query: 61 AINPTPAGEGKSTMSIGLADALNKGKKTMIADREPSLGPVMGIKGAAGGGYAQVLPME 120
 AINPTPAGEGKST++IGLADALNKGKKTMIADREPSLGPVMGIKGAAGGGYAQVLPME
 Sbjct: 61 AINPTPAGEGKSTITIGLADALNKGKKTMIADREPSLGPVMGIKGAAGGGYAQVLPME 120

Query: 121 DINLHFTGDMHAITTANNALSALLDNHILHOGNELDIDQRRVIWKRVDLNDRLRQVIVG 180
 DINLHFTGDMHAITTANNALSAL+DNH+HOGNEL IDQRR+IWKRVDLNDRLR V VG
 Sbjct: 121 DINLHFTGDMHAITTANNALSALIDNHILHOGNELGIDQRRIIWKRVDLNDRLRHVTVG 180

Query: 181 LGSPVNGIPREDGFDITVASEIMAILCLATDLSLKKRLSNIVVAYSRRNKPIYVKDLKI 240
 LGSP+NGIPREDGFDITVASEIMAILCLAT++ DLK+RL+NIV+ Y +R P+YV+DL++
 Sbjct: 181 LGSPINGIPREDGFDITVASEIMAILCLATNVEDLKERLANIVIGYRFDSPVYVRDLEV 240

Query: 241 EGALTILKDTIKPNLVQTIYGTTPALVHGPPFANIAHGCVSLATSTALRLADYVVT EAG 300
 +GAL LILK+ IKPNLVQTIYGTTPA VHGGPPFANIAHGCVSLATSTALRLADY +TEAG

10

15

20

25

30

The protein has homology with the following sequences in the databases:

35

40

45

50

55

60

65

Query: 361 KRHVENMRQFKVPVVVAINEFVADTEAEIATLTKALCEEIKVPVELASVWANGEGGLALA 420
RHVENMR++ VPVVVAINEF+ DT EIA L+ LC I VPVELASVWANGA+GG+ LA
Sbjct: 361 ARHVENMRKYGVPVVVAINEFITDND EIAVLRLNLC AIDVPVELASVWANGADGGVDLA 420

-2253-

5 Query: 421 KTVVRVIDQEADYKRLYSDEDTLEEKVINIVTQIYGGKAVQFGPKAKTQLKQFAEFGWD 480
 T++ I+ + YKRLY + ++EEKV I +IY V F KAKTQ+ Q + GWD
 Sbjct: 421 NTLINTIENNPNPHYKRLYDNNLSVEEKVTEIAKETIRADKVIPEKKAKTQIAQIVKNGWD 480

Query: 481 KLPVCMAKTQYSFSDNPSLLGAPTFDITIREFVPKTGAGFIVGLTGDVMTMPGLPKVPA 540
 LP+CMAKTQYSFSD+P LLGAPT FDIITIRE VPK GAGFIV LTGDVMTMPGLPK PA
 Sbjct: 481 NLPICMAKTQYSFSDDPKLLGAPTFDITIRELVPKLGAGFIVALTGDVMTMPGLPKKPA 540

10 Query: 541 AMAMDVAENG TALGLF 556
 A+ MDVA +GTALGLF
 Sbjct: 541 ALNMDVAADGTALGLF 556

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 452/556 (81%), Positives = 513/556 (91%)

Query: 1 MKTDIEIAQSVALKPTAEIVEQVGIGFDDIELYGKYKAKLSFDKIEAVKSQKVGLILVT 60
 MK+DIEIAQSVAL+PI +IV++VGI DDIELYGKYKAKLSF+K++AV++ + GKLILVT
 20 Sbjct: 1 MKSDIEIAQSVALQPTITDIVKKVIGDGGDIELYGKYKAKLSFEKMKAVEANEPGKLILVT 60

Query: 61 AINPTPAGEGKSTMSIGLADALNKIGKKTMIALREPSLGPVMGIKGAAGGGYAQVLPME 120
 AINPTPAGEGKSTMSIGLADALN++GKKTMI+ALREPSLGPVMGIKGAAGGGYAQVLPME
 Sbjct: 61 AINPTPAGEGKSTMSIGLADALNQMGKKTMLALREPSLGPVMGIKGAAGGGYAQVLPME 120

25 Query: 121 DINLHFTGDMHAITTANNALSALLDNHIIHQGNELDIDQRRVIWKRVDLNDRLRQVIVG 180
 DINLHFTGDMHAITTANNALSAL+DNH+ QGN+L ID RR+IWKRV+DLNDRLRQVIVG
 Sbjct: 121 DINLHFTGDMHAITTANNALSALIDNHLQQGNDLGIDPRRIIWKRVLDLNDRLRQVIVG 180

30 Query: 181 LGSPVNGIPREDGFDTIVASEIMAILCLATDLSDLKKRLSNIVVAYSRRNKPIYVKDLKI 240
 LGSPVNG+PREDFDTIVASEIMAILCLATDL DLKKRL++IVVAY+ +RKP+YV+DLK+
 Sbjct: 181 LGSPVNGVPREDFDTIVASEIMAILCLATDLKDLKKRLADIVVAYTYDRKPVYVRDLKV 240

35 Query: 241 EGALTILIKDITIKPNLVQTIYGTALVHGGPFANIAHGCNSVLATSTALRLADYVTEAG 300
 EGALTILIKD IKPNLVQTIYGTALVHGGPFANIAHGCNSVLATSTALRLADY VTEAG
 Sbjct: 241 EGALTILIKDAIKPNLVQTIYGTALVHGGPFANIAHGCNSVLATSTALRLADYTVTEAG 300

40 Query: 301 FGADLGAEKFLDIKTPNLPTSPDAIVIVATLRALKMHGGVSKEDLSQENVEAVKRGFTNL 360
 FGADLGAEKFL+IK PNLPT+PDAIVIVATLRALKMHGGV+K DL+ EN EAV+ GF NL
 Sbjct: 301 FGADLGAEKFLNIKVPNLTPKAPDAIVIVATLRALKMHGGVAKSDLAENCEAVRLGFPANL 360

45 Query: 361 ERHVNMRQYGVVPPVVAINQFTADTESEIATLTKLCSNIDVAVELASVWEDGADGGLLELA 420
 +RHV NMRQ+ VPVVVAIN+F ADTE+EIATLK LC I V VELASVW +GA+GGL LA
 Sbjct: 361 KRHVENMRQFKVPVVVAINNEFVADTEAEIATLKALCEEIKVPVELASVWANGAEGGLALA 420

50 Query: 421 QTVANVIETQSSNYKRLYNDEDTIEEKIKKIVTKIYGGNKVHFGPKAQIQLKEFSNDNGWD 480
 +TV VI+ +++YKRLY+DEDT+EEK+ IVT+IYGG V FGPKA+ QLK+F++ GWD
 Sbjct: 421 KTVVRVIDQEADYKRLYSDEDTLEEKVINIVTQIYGGKAVQFGPKAKTQLKQFAEFGWD 480

Query: 481 KMPICMAKTQYSFSDNPNLLGAPTFDITVREFVPKTGAGFIVALTGDVLTMPGLPKKPA 540
 K+P+CMAKTQYSFSDNP+LLGAPTFDIT+REFVPKTGAGFIV LTGDV+TMPGLPK PA
 Sbjct: 481 KLPVCMAKTQYSFSDNPSLLGAPTFDITIREFVPKTGAGFIVGLTGDVMTMPGLPKVPA 540

55 Query: 541 ALNMDVLEDGTAIGLF 556
 A+ MDV E+GTA+GLF
 Sbjct: 541 AMAMDVAENG TALGLF 556

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 9057> which encodes amino acid sequence <SEQ ID 9058>. Analysis of this protein sequence reveals the following:

60 Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.49 Transmembrane 516 - 532 (516 - 533)

----- Final Results -----
 bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2254-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS sequences follows:

Score = 604 bits (1540), Expect = e-174
 Identities = 304/555 (54%), Positives = 389/555 (69%), Gaps = 2/555 (0%)

Query: 4 SDIEIANSVTMEFISKVADQLGIDKEALCLYGKYKAKIDARQLVALKNKPDGKLILVTAI 63
 +DIEIA SV ++PI+++ +Q+GI + + LYGKYKAK+ ++ A+K++ GKLILVTAI
 Sbjct: 3 TDIEIAQSVALKPIAEIVEQVGIGFDDIELYGKYKAKLSFDKIEAVKSQKVGLILVTAI 62

Query: 64 SPTPAGEGKTTTSVGLVDALSAGKKAVIALREPSLXXXXXXXXXXXXXXXXXXPMEDI 123
 +PTPAGEGK+T S+GL DAL+ IGKK +IALREPSL PMEDI
 Sbjct: 63 NPTPAGEGKSTMSIGLADALNKIGKKTMIALREPSLGPVMGIKGAAGGGYAQVLPMDI 122

Query: 124 NLHFTGDFHAIQVANNLLAALIDNHHHNSLGLDSRRITWKRVDMMNDRLRHIVDGLQ 183
 NLHFTGD HAI ANN L+AL+DNHH GN L ID RR+ WKRVD+NDR LR ++ GL
 Sbjct: 123 NLHFTGDMHAITANNALSALLDNHHHNSLGLDSRRITWKRVDMMNDRLRHIVDGLQ 182

Query: 184 GKVNGIPREDGYDITVASEIMAILCLSENISDLKARLEKIIIGVNYQGEFVTXXXXXXX 243
 VNGIPREDG+DITVASEIMAILCL+ ++SDLK RL I++ Y+ +P+
 Sbjct: 183 SPVNGIPREDGFDITVASEIMAILCLATDLSDLKKRLSNIVVAYSRRNKPIYVKDLKIEG 242

Query: 244 XXXXXXXXXIHFNLVQTLHTPALIHGGPFANIAHGCNSVLATKLALKYGDYAVTEAGFG 303
 I PNLVQT+ TPAL+HGGPFANIAHGCNSVLAT AL+ DY VTEAGFG
 Sbjct: 243 ALTLLIKDTIKPNLVQTIYGPALVHGGPFANIAHGCNSVLATSTALRLADYVTEAGFG 302

Query: 304 ADLGAKEFIDIKRMSGRLPAAVVLVATIRALKMHGGVPKADLATENVQAVVDGLPNLDK 363
 ADLGAKEF+DIK P A+V+VAT+RALKMGGV K DL+ ENV+AV G NL++
 Sbjct: 303 ADLGAKEFLDIKTPNLPTSPDAIVIVATLRLALKMHGGVSKEDLSQENVEAVKRGFTNLER 362

Query: 364 HLANIQDVYGLPVVVAINKFPLDTELQAVYDACDKRGVDVVISDVWANGAGGRELAE 423
 H+ N++ YG+PVVVAIN+F DT++E+ + C V V ++ VW +G GG ELA+
 Sbjct: 363 HVNMMRQ-YGVPVVVAINQFTADTESEIATLTLCSNIDVAELASVWEDGADGGLLELAQ 421

Query: 424 KVVTLAE-QDNQFRFVYEEEDSIEITKLTIVTKVYGGKGINLSSAAKRELADLERLGFGN 482
 V + E Q + ++ +Y ++D+IE K+ KIVTK+YGG ++ A+ +L + G+
 Sbjct: 422 TVANVIETQSSNYKRLYNDEDTIEEKIKKIVTKIYGNKVFHFGPKAQIQLKEFSDNGWDK 481

Query: 483 YPICMAKTQYSFSDDAKKLGAPTDFVTISNLKVSAGAGFIVALTGAITMPGLPKVPAS 542
 PICMAKTQYSFSD+ LGAPTDF +T+ GAGFIVALTG ++TMPGLPK PA+
 Sbjct: 482 MPICMAKTQYSFSDNPNLLGAPTDFDITVREFVPKTGAGFIVALTGDVLTMPGLPKKPAA 541

Query: 543 ETIDIDEEGNITGLF 557
 +D+ E+G GLF
 Sbjct: 542 LNMDVLEDGTAIGLF 556

SEQ ID 6198 (GBS131) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 6; MW 64.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 4; MW 90kDa).

GBS131-GST was purified as shown in Figure 201, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2005

A DNA sequence (GBSx2115) was identified in *S.agalactiae* <SEQ ID 6201> which encodes the amino acid sequence <SEQ ID 6202>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood =-10.03 Transmembrane 34 - 50 (29 - 56)

-2255-

INTEGRAL	Likelihood = -7.70	Transmembrane	90 - 106 (84 - 110)
INTEGRAL	Likelihood = -1.97	Transmembrane	62 - 78 (62 - 78)
INTEGRAL	Likelihood = -0.69	Transmembrane	275 - 291 (275 - 291)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA88609 GB:M37842 unknown protein [Streptococcus mutans]
 Identities = 243/373 (65%), Positives = 302/373 (80%), Gaps = 1/373 (0%)

Query: 71 IGAVLYLVNSEMDALSRVTWLILVMIAPLLGAMFLMYTKFDWGYRGLKQRLLETLLIDESQI 130
 15 IG+VLYLVNS+MD LS +TWL++++ P+LG +FL+YTK DWGYR LK ++ +
 Sbjct: 2 IGSVLYLVNSQMDTSLIITWLLVILPFPILGTFLIYTKQDWGYRELKSLIKKSTQAIKP 61

Query: 131 YLEDDPETLNQLKSSTSTTYHLVQYFEKAHGNFVYRNTDVTFLPTGEAFFEKMKEELLK 190
 Y + D L +LK S + TY+L QY ++ G FVY+NT VT+ P G++ FE+MK++LLK
 20 Sbjct: 62 YFQYDQRILYKLKESHARTYNLAQYLHRS-GGFPVYKNTKVYTFPNGQSKFEEMKKQLLK 120

Query: 191 AKKYIFLEFFIIDEGIMWGEILSILEQKVEEGVEVRILYDGMIEITKLSFDYTKRLEKIG 250
 A+K+IFLE+FII EG+MWGEILSILEQKV+EGVEVR++YDGM+E++ LSFYD KRLEKIG
 25 Sbjct: 121 AEKFIFLEYFIIAEGLMWGEILSILEQKVEEGVEVRVMYDGMLELSTLSFDYAKRLEKIG 180

Query: 251 IKAKAFSPISPFISTYNYNRDHRKIVVIDGVVGMTGGVNLADHEYINHIEFGHWKDSGIM 310
 IKAK FSPI+PF+STYNYNRDHRKI+VID V GG+NLADHEYIN IE FG+WKD+ +M
 30 Sbjct: 181 IKAKVFSPI+PFVSTYNYNRDHRKILVIDNKVAFNGGINLADEYINQIERFGYWKDTAVM 240

Query: 311 LKGKAVDSFLLLFLQMWSTTEKMLVAPYLGVDLLENEGYVIPYGDSPDLTDKVGENV 370
 L+G+ V SF L+FLQMWST + APYL + + GYVIPY DSPLD +KVGENV
 35 Sbjct: 241 LEGEGVASFTLMFLQMWSTTNKDYEFAPYLTQNFHEIVANGYVIPYSDSPLDHEKVGENV 300

Query: 371 YIDILNHAREYVYIMTPYLILDSELEHAIQFAAERGVDVRIIMPGIPDKPIPYALAKTY 430
 YIDILN AR+YVYIMTPYLILDSE+EHA+QFAAERGVDV+IIMPGIPDK +P+ALAK Y+
 40 Sbjct: 301 YIDILNQARDYVYIMTPYLILDSEMEHALQFAAERGVDVKIIMPGIPDKKVPFALAKRYF 360

Query: 431 QALTSGSVKIYEY 443
 AL +GVKIYE+
 Sbjct: 361 PALLDAGVKIYEF 373

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6203> which encodes the amino acid sequence <SEQ ID 6204>. Analysis of this protein sequence reveals the following:

Possible site: 47
 45 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.86	Transmembrane	84 - 100 (81 - 104)
INTEGRAL	Likelihood = -8.33	Transmembrane	28 - 44 (23 - 49)
INTEGRAL	Likelihood = -6.74	Transmembrane	56 - 72 (53 - 74)

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 The protein has homology with the following sequences in the databases:

>GP:AAA23240 GB:J02911 formyltetrahydrofolate synthetase (FTHFS)
 (ttg start codon) (EC 6.3.4.3) [Moorella thermoacetica]
 Identities = 350/557 (62%), Positives = 438/557 (77%), Gaps = 2/557 (0%)

Query: 2 VLSDIEIANSVTMEPIISKVADQLGIDKEALCLYGYKAKIDARQLVALKNKPDGKLILVT 61
 V SDIEIA + M+P+ ++A LGI ++ + LYGYKAKI LK+KPDGKLILVT
 60 Sbjct: 4 VPSDIEIAQAAMKPMELARGLGIQDEVELYGYKAKISLDVYRRLKDKPDGKLILVT 63

Query: 62 AISPTPAGEGKTTTSVGLVDALSAGKKAVIALREPSLGPVFGVKGGAAGGHAQVVPME 121

-2256-

AI+PTPAGEGKTTTSVGL DAL+ +GK+ ++ LREPSLGP FG+KGAAGGG+AQVVPME
 Sbjct: 64 AITPTPAGEGKTTTSVGLTDALARLQKRVMLCLREPSLGPSPGKGAAGGGYAQVVPME 123

Query: 122 DINLHFTGDFHAIGVANNLLAALIDNHHHNSLIGDSRRITWKRVDMMNDROLRHIVDG 181
 DINLHFTGD HA+ A+NLAA++DNH+ GN L ID R ITW+RV+D+NDR LR+IV G
 Sbjct: 124 DINLHFTGDIHAVTYAHNLLAAMVDNHLQQGNVLNIDPRTITWRRVIDLNDRLRNIVIG 183

Query: 182 LQGVNIGIPREDGYDITVASEIMAILCLSENISDLKARLEKIIIGYNYQGEFVTAADLKA 241
 L GK NG+PRE G+DI+VASE+MA LCL+ ++ DLK R +I++GY Y G+PVT A DL+A
 Sbjct: 184 LGGKANGVPRETGFDISVASEVMACCLASDLMDLKERFSRIVVGTYDGKPVTAAGDLEA 243

Query: 242 GGALAALLKDAIHPNLVQTLLEHTPALIHGGPFANIAHGCNSVLATKLALKYGDYAVTEAG 301
 G++A L+KDAI PNLVQTLLE+TPA IHGGPFANIAHGCNS++ATK ALK DY VTEAG
 Sbjct: 244 QGSMALLMKDAIKPNLVQTLLENTPAFIHGGPFANIAHGCNSIIATKTALKLADYVVTEAG 303

Query: 302 FGADLGAEKFDIDIKCRMSGSLRPAAVVLVATIRALKMHGGVPKADLATENVQAVDGLPNL 361
 FGADLGAEKF D+KCR +G +P A V+VAT+RALKMHGGVPK+DLATEN++A+ +G NL
 Sbjct: 304 FGADLGAEKFYDVCKRYAGFKPDATVIVATVRALKMHGGVPKSDLATENLEALREGFANL 363

Query: 362 DKHLANIQDVYGLPVVVAINKFPLDTEAELQAVYDACDKRGVDVVISDVWANGGAGGREL 421
 +KH+ NI +G+P VVAIN FP DT+AEL +Y+ C K G +V +S+VWA GG GG EL
 Sbjct: 364 EKHIENI-GKFGVPAVVAINAFPTDTEAELNLLYELCAKAGAEVALSEVWAKGGGGLEL 422

Query: 422 AEKV- TLAEQDNQFRFVYEDDSIETKLTIVTKVYGGKGINLSSAAKRELADLERLGF 480
 A KV+ TL + + F +Y D SI+ K+ KI T++YG G+N ++ A + + E LG+
 Sbjct: 423 ARKVLQTLSESRPSNFHVLNLDLSIKDKIAKATEIYGADGVNNTYAEADKAIQRYESLGY 482

Query: 481 GNYPICMAKTQYSFSDDAKKGAPTDFTVTISNLKVSAGAGFIVALTGAIMTMPGLPKVP 540
 GN P+ MAKTQYSFSD KLG P +FT+T+ +++SAG IV +TGAIMTMPGLPK P
 Sbjct: 483 GNLPVVMAKTQYSFSDDMTKLGRPRNFTITVREVRLSAGGRILVPTGAIMTMPGLPKRP 542

Query: 541 ASETIDIDEENITGLF 557
 A+ IDID +G ITGLF
 Sbjct: 543 AACNIDIDADGVITGLF 559

!GB:M37842 unknown protein [Streptococcus mutans] (v... 517 e-145
 >GP:AAA88609 GB:M37842 unknown protein [Streptococcus mutans]
 Identities = 246/370 (66%), Positives = 303/370 (81%), Gaps = 1/370 (0%)

Query: 68 VLYLVNSMDAISRMFWLILIMIAPLGLSLFLIYTKLDWGYRGLKQRINHLVDLSAPYLS 127
 VLYLVNS MD +S +TWL++I+ P+LG+LFLIYTK DWGYR LK I PY
 Sbjct: 5 VLYLVNSQMDTSLITLWLVILPFPILGLTFLIYTKQDWGYRELKSLIKKSTQAIKPYFQ 64

Query: 128 DDDAILEVLKDSSTTYHLVQYRLERSRGNFPIYNNTRVITYFPTGETFFDSLKEQLFLAKK 187
 D IL LK+S + TY+L QYL RS G FP+Y NT+VTYFP G++ F+ +K+QL A+K
 Sbjct: 65 YDQRLYKLKESHARTYNLAQYLHRS-GGFPVYKNTKVITYFPNGQSKFEEMKKQLLKAEK 123

Query: 188 YIFLEFFIIAEGQMWGEILSILEKKVSEGEVVRVLFDMNELSTLSSDYAKRLEQIGIKA 247
 +IFLE+FIIAEG MWGEILSILE+KV EGVEVRV++DGM ELSTLS DYAKRLE+IGIKA
 Sbjct: 124 FIFLEYFIIAEGLMWGEILSILEQKQVEGEVVRVMDGMLELSTLSFDYAKRLEKIGIKA 183

Query: 248 KSFLPISPFISTYYNYRDHRKIVVIDGEVSFTGGINLADEYINEVERFGHWKDAGLMLEG 307
 K F PI+PF+STYYNYRDHRKI+VID +V+F GGINLADEYIN++ERFG+WKD +MLEG
 Sbjct: 184 KVFSPIPFVSTYYNYRDHRKILVIDNKVAFNGGINLADEYINQIERFGYWKDTAVMLEG 243

Query: 308 EATDSFLILFLQMWSTITEKELIIDPYLSDHSLKLPDGYVPIPYGDSPLDTPDKIGNVYID 367
 E SF ++FLQWWS T K+ PYL+ + ++ ++GYVPIPY DSEPLD +K+G+NVYID
 Sbjct: 244 EGVASFTLMFLQMWSTINKDYEFAPYLTQNFHEIVANGYVPIPYSDSPLDHEKVGENVYID 303

Query: 368 ILNHAKEYVYIMTPYLILDSEMEHALRFASERGVDIRIIMPVDPKGVFPYALAKTYKAL 427
 ILN A++VYVIMTPYLILDSEMEHAL+FA+ERGVD++IIMP+PDK VP+ALAK Y+ AL
 Sbjct: 304 ILNQARDYVYIMTPYLILDSEMEHALQFAAERGVDVKIIMPVDPKGVFPYALAKRYFPAL 363

Query: 428 MSSGVKIYEF 437
 + +GVKIYEF
 Sbjct: 364 LDAGVKIYEF 373

An alignment of the GAS and GBS proteins is shown below.

-2257-

Identities = 362/524 (69%), Positives = 437/524 (83%)

Query: 8 LISNKVKIVRLNKSLLRGIFSRRTTIVAILLILQLLFLASYSWLEQYRVWLATVEH 67
 +I K K+ LL+K K LRGIFSRRTT+I +L+ILQL+FL SY+W+EQYRVW+ +E
 5 Sbjet: 2 IIKKAKVKYLLHKHKGFLRGIFSRRTTIIIVLLILQLVFLFQSYAWMEQYRVWITILES 61

Query: 68 ILTIGAVLYLVNSEMDALSRVTWLILVMIAPLIGAMFLMYTKFDWGYRGLKQRLETLIDE 127
 + I VLYLVNS+MDA+SR+TWLIL+MIAPLLG++FL+YTK DWGYRGLKQR+ L+D
 10 Sbjet: 62 VFATITVLYLVNSMDAISRMTWLILIMIAPLLGSFLIYTKLDWGYRGLKQRINHLVDL 121

Query: 128 SQIYLEDDEPETLNQLKSSTSTTYHLVQYFEKAHGNFPVYRNTDVTFLPTGEAFFEKMKEE 187
 S YL DD L LK STSTTYHLVQY E++ GNFP+Y NT VT+ PTGE FF+ +KE+
 15 Sbjet: 122 SAPYLSDDDAILEVLKDSTSTTYHLVQYLERSRGNFPYNNTRVITYFPTGETFFDSLKEQ 181

Query: 188 LLKAKKYIFLEFFIIDEGIMWGEILSILEQKVBEQVEVRILYDGMIEITKLSFDYTKRLE 247
 L AKKYIFLEFFII EG MWGEILSILE+KV EGVEVR+L+DGM E++ LS DY KRLE
 20 Sbjet: 182 LFLAKKYIFLEFFIIAEGQMWGEILSILEKKVSEGVVEVRVLFDMNELSTLSSDYAKRLE 241

Query: 248 KIGIKAKAFSPISPFISTYYNYRDHRKIVVIDGVVGMTGCVNLADEXINHIELFGHWKDS 307
 +IGIKAK+F PISPFISTYYNYRDHRKIVVIDG V TGG+NLADEXIN +E FGHWKD+
 25 Sbjet: 242 QIGIKAKSFLPISPFISTYYNYRDHRKIVVIDGEVSFTGGINLADEXINEVERFGHWKDA 301

Query: 308 GIMLKGAVDLSFLLFLQMWSITEEKMLVAPYLGVDHDLVENEGYVIPYGDSPDLTDKVG 367
 G+ML+G+A DSFL+LFLQMWSITE+++++ PYL H + ++GYVIPYGDSPDLTDK+G
 30 Sbjet: 302 GLMLEGEATDSFLILFLQMWSITEKELIIDPYLSDHSLKLPDGYVIPYGDSPDLTDKIG 361

Query: 368 ENVYIDILNHAREYVYIMTPYLILDSELEHAIQFAAERGVDVRIIMPPIPDKPIPYALAK 427
 +NVYIDILNHA+EYVYIMTPYLILDSE+EHA++FA+ERGVD+RIIMP+PDK +PYALAK
 35 Sbjet: 362 KNVYIDILNHAKEYVYIMTPYLILDSEMEHALRFASERGVDIRIIMPGVDPKGVYALAK 421

Query: 428 TYYQALTKSGVKIYEYTLGFVHSKIFLSDNTKAVVGTINLDYRSLYHHFECAVLYKVDA 487
 TYY+AL SGVKIYEY GFVHSK+F+SDNTKAVVGTINLDYRSLYHHFECA YLY+V
 40 Sbjet: 422 TYYKALMSSGVKIYEYQPGFVHSKVFI SDNTKAVVGTINLDYRSLYHHFECAVLYRVSV 481

Query: 488 IQDIYRDYMDTLNKSRLVSLKDINNIPKFQKVIGIVTKTIAPLL 531
 I DI D+ + +S L++ + P +QK+IG++ + IAPLL
 45 Sbjet: 482 IADIVNDFNEAQKQSLMTSDHLTQRPWYQKLIGLLVRIIAPLL 525

40 A related GBS gene <SEQ ID 8953> and protein <SEQ ID 8954> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 6
 McG: Discrim Score: -8.80
 GvH: Signal Score (-7.5): -1.94
 Possible site: 53
 45 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 4 value: -10.03 threshold: 0.0
 INTEGRAL Likelihood = -10.03 Transmembrane 34 - 50 (29 - 56)
 INTEGRAL Likelihood = -7.70 Transmembrane 90 - 106 (84 - 110)
 INTEGRAL Likelihood = -1.97 Transmembrane 62 - 78 (62 - 78)
 50 PERIPHERAL Likelihood = 1.22 199
 modified ALOM score: 2.51
 *** Reasoning Step: 3
 55 ----- Final Results -----
 bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

60 The protein has homology with the following sequences in the databases:

32.5/57.2% over 498aa
 Bacillus firmus
 SP|066043| CARDIOLIPIN SYNTHETASE (EC 2.7.8.-) (CARDIOLIPIN SYNTHASE) (CL SYNTHASE).
 Insert characterized
 65 GP|2952028|gb|AAC05444.1||U88888 cardiolipin synthase Insert characterized

ORF01572 (409 - 1893 of 2193)
 SP|066043|CLS_BACFI (5 - 503 of 503) CARDIOLIPIN SYNTHETASE (EC 2.7.8.-) (CARDIOLIPIN
 SYNTHASE) (CL SYNTHASE). GP|2952028|gb|AAC05444.1|U88888 cardiolipin synthase {Bacillus
 firmus}
 %Match = 17.9
 %Identity = 32.5 %Similarity = 57.1
 Matches = 162 Mismatches = 204 Conservative Sub.s = 123

55 SEQ ID 8954 (GBS277d) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 150 (lane 18; MW 51kDa), in Figure 151 (lane 17 & 18; MW 51kDa) and in Figure 182 (lane 12; MW 51kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 151 (lane 15 & 16; MW 76kDa) and in Figure 58 (lane 5; MW 87kDa).

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2259-

Example 2006

A DNA sequence (GBSx2116) was identified in *S.agalactiae* <SEQ ID 6205> which encodes the amino acid sequence <SEQ ID 6206>. This protein is predicted to be aspartate-semialdehyde dehydrogenase. Analysis of this protein sequence reveals the following:

```

5      Possible site: 42
      >>> Seems to have an uncleavable N-term signal seq

      ----- Final Results -----
10      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9831> which encodes amino acid sequence <SEQ ID 9832>
was also identified.

15  The protein has homology with the following sequences in the GENPEPT database.

      >GP:AAA26850 GB:J02667 aspartate beta-semialdehyde dehydrogenase (EC
          1.2.1.11) [Streptococcus mutans]
          Identities = 261/357 (73%), Positives = 304/357 (85%), Gaps = 1/357 (0%)

20  Query: 1  MGYTVAIVGATGAVGTQMIRQLEQSNLPTEQVKLLSSSSRSAGKILHFKDEAIRVEETKE 60
      MGYTVAIVGATGAVGT+MI+QLEQS LP+++V+LLSSSSRSAGK+L +KD+ + VE TTK+
      Sbjct: 1  MGYTVAIVGATGAVGTRMIQQLEQSTLPVDKVRLLSSSSRSAGKVLQYKDQDVTVELTTKD 60

25  Query: 61  SFYDVDIALFSAGGSISAKFAPYAVKSGAVVDNTSYFRQNPDPVPLVPEVNAHAMIGHN 120
      SF VDIALFSAGGS+SAKFAPYAVK+GAVVDNTS+FRQNPDPVPLVPEVNA+AM HN
      Sbjct: 61  SFEAVDIALFSAGGSVSASFAPYAVKAGAVVDNTSHFRQNPDPVPLVPEVNAYAMDAHN 120

      Query: 121 GIIACPN CSTIQMMIALEPIRQKWGIERVIVSTYQAVSGSGARAVEETKEQLRQVLNDNL 180
      GIIACPN CSTIQMM+ALEPIRQKWG+ RVIVSTYQAVSG+G A+ ET ++++V+ND +
30  Sbjct: 121 GIIACPN CSTIQMMVALEPIRQKWGLSRVIVSTYQAVSGAGQSAINETVREIKEVVDGVL 180

      Query: 181 SPDQLIATVLPSSDQKHYPFALPQIDIFTDNDYTYEEMKMTLETKKIMEDATIKVS 240
      P + A + P D+KHYPIAFNAL QID+FTDNDYTYEEMKMT ETKKIME+ + VS
35  Sbjct: 181 DPKAVHADIFPSGGDKKHYPIAFNALAQIDVFTDNDYTYEEMKMTNETKKIMEEPELPVS 240

      Query: 241 ATCVIRIPVLSGHSESIYIETKELASISEIKKAIANFPGAVLQDLPSQQIYPQAINAVGHR 300
      A CVR+P+L HSE++YIETK++A I E+K AIA FPGAVL+D QIYPQA NAVG R
      Sbjct: 241 AHCVRVPILFSHSEAVYIETKDVAPIEEVKAAIAAFPGAVLEDDIKHQIYPQAA NAVGSR 300

40  Query: 301 ETFVGRIRKDLQENGVMHVVSDNLLKGAAWN SVQIAETLHKNGLVKPAKELKFEL 357
      ETFVGRIRKDLQ ENG+HMMVVSDNLLKGAAWN+ A LH+ GLV+ ELKFEL
      Sbjct: 301 -ETVGRIRKDLQENGVMHVVSDNLLKGAAWN SIITANRLHERGLVRSTSELKFEL 356

```

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2007

A DNA sequence (GBSx2117) was identified in *S.agalactiae* <SEQ ID 6207> which encodes the amino acid sequence <SEQ ID 6208>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 22
      >>> Seems to have no N-terminal signal sequence
          INTEGRAL Likelihood = -2.66 Transmembrane 33 - 49 ( 33 - 49)

      ----- Final Results -----
55  bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-2260-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 500.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2008

A DNA sequence (GBSx2119) was identified in *S.agalactiae* <SEQ ID 6209> which encodes the amino acid sequence <SEQ ID 6210>. Analysis of this protein sequence reveals the following:

10 Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.3853 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2009

- A DNA sequence (GBSx2120) was identified in *S.agalactiae* <SEQ ID 6211> which encodes the amino acid sequence <SEQ ID 6212>. This protein is predicted to be unnamed protein product (clpP). Analysis of this protein sequence reveals the following:

25 Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.3883 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 35 A related GBS nucleic acid sequence <SEQ ID 10061> which encodes amino acid sequence <SEQ ID 10062> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6213> which encodes the amino acid sequence <SEQ ID 6214>. Analysis of this protein sequence reveals the following:

40 Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2682 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 175/196 (89%), Positives = 187/196 (95%)

-2261-

Query: 5 MIPVVIEQTSRGERSDIYSRLKDRIMLTGQVEDNMANSIIAQLFLDAQDNTKDIYL 64
 MIPVVIEQTSRGERSDIYSRLKDRIMLTG VEDNMANS+IAQLFLDAQDNTKDIYL
 Sbjet: 1 MIPVVIEQTSRGERSDIYSRLKDRIMLTGPFVEDNMANSVIAQLFLDAQDNTKDIYL 60

5 Query: 65 YVNTPGGSVSAGLAIVDTMNFNIKSDVQTIVMGMAASMGTTIASSGAKGKRFMLPNAEYMI 124
 YVNTPGGSVSAGLAIVDTMNFNIK+DVQTIVMGMAASMGTT+IASSG KGKRFMLPNAEYMI
 Sbjet: 61 YVNTPGGSVSAGLAIVDTMNFNIKADVQTIVMGMAASMGTTIASSGKTGKRFMLPNAEYMI 120

10 Query: 125 HQPMGGTGGGTQQSDMAIAAEHLKTRHTLEKILADNSGQSIEKVHDDAERDRWMSAQET 184
 HQPMGGTGGGTQQ+DMAIAAEHLKTRH LEKILA N+G++I+++H DAERD WMSA+ET
 Sbjet: 121 HQPMGGTGGGTQQT'DMAIAAEHLKTRHLEKILAQNAGKTIKQIHKDAERDYWMSAEET 180

Query: 185 LDYGFIDAIMENNNLQ 200
 L YGFID IMENN L+

15 Sbjet: 181 LAYGFIDEIMENNELK 196

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2010

20 A DNA sequence (GBSx2121) was identified in *S.agalactiae* <SEQ ID 6215> which encodes the amino acid sequence <SEQ ID 6216>. This protein is predicted to be uracil phosphoribosyltransferase (upp). Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

25 INTEGRAL Likelihood = -0.43 Transmembrane 127 - 143 (127 - 144)
 INTEGRAL Likelihood = -0.06 Transmembrane 72 - 88 (72 - 89)
 INTEGRAL Likelihood = -0.06 Transmembrane 154 - 170 (154 - 170)

30 ----- Final Results -----
 bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35 A related GBS nucleic acid sequence <SEQ ID 10063> which encodes amino acid sequence <SEQ ID 10064> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA26890 GB:L07793 uracil phosphoribosyltransferase
 [Streptococcus salivarius]
 Identities = 192/209 (91%), Positives = 202/209 (95%)

40 Query: 1 MGKFQVISHPLIQHKLILRRTTTSTKDFRELVDIAMLGMGYEVSRLDPLEDVEIQT'PVA 60
 MGKFQVISHPLIQHKLILRR TSTKDFRELVDIAMLGMGYEVSRLDPLE+VEIQT'+
 Sbjet: 1 MGKFQVISHPLIQHKLILRRDSTKDFRELVDIAMLGMGYEVSRLDPLEEVEIQT'PIT 60

45 Query: 61 TTVQKQLAGKKLAIVPILRAGIGMVDGFLSLVPAKVGHIGMYRDEETFQPV EYLVKLPE 120
 TVQKQL+GKKLAIVPILRAGIGMVDGFLSLVPAKVGHIGMYRDEET +PVEYLVKLPE
 Sbjet: 61 KTVQKQLSGKKLAIVPILRAGIGMVDGFLSLVPAKVGHIGMYRDEETLEPVEYLVKLPE 120

50 Query: 121 DIDQRQIFVVDPLATGGSAILAVDSLKKRGAASIKFVCLVAAPEGVAALQEAHPD'VDIY 180
 DIDQRQIFVVDPLATGGSAILAVDSLKKRGAA+IKFVCLVAAPEGV LQ+AHPD+DIY
 Sbjet: 121 DIDQRQIFVVDPLATGGSAILAVDSLKKRGAANIKFVCLVAAPEGVKQLQDAHPDID'Y 180

Query: 181 TAALDEKLNEHGYYVPGLDAGDRLFGTK 209
 TA+LDEKLNE+GYIVPGLDAGDRLFGTK

55 Sbjet: 181 TASLDEKLNEGYIVPGLDAGDRLFGTK 209

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6217> which encodes the amino acid sequence <SEQ ID 6218>. Analysis of this protein sequence reveals the following:

-2262-

Possible site: 26
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.59 Transmembrane 72 - 88 (72 - 89)
 INTEGRAL Likelihood = -0.22 Transmembrane 127 - 143 (127 - 144)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein is similar to uracil phosphoribosyltransferase from *S.salivarius*:

>GP:AAA26890 GB:L07793 uracil phosphoribosyltransferase [Streptococcus salivarius]
 Identities = 191/209 (91%), Positives = 205/209 (97%)

Query: 1 MGKQVISHPLIQHKLSILRRQTTSTKDFRELVNEIAMLMGYEVSRLPLEDVDIQTTPVS 60
 MGK QVISHPLIQHKLSILRR+ TSTKDFRELVNEIAMLMGYEVSRLPLE+V+IQTP++
 Sbjct: 1 MGKQVISHPLIQHKLSILRREDTSTKDFRELVNEIAMLMGYEVSRLPLEEVEIQTPIT 60
 Query: 61 KTVQKQLAGKKLAIVPILRAGIGMVDGLLSLVPAAKVGHIGMYRNEETLEPVEYLVKLPE 120
 KTVQKQL+GKKLAIVPILRAGIGMVDG LSLVPAAKVGHIGMYR+EETLEPVEYLVKLPE
 Sbjct: 61 KTVQKQLSGKKLAIVPILRAGIGMVDGFLSLVPAAKVGHIGMYRDEETLEPVEYLVKLPE 120
 Query: 121 DINQRQIFLVDPMPLATGGSAILAVDSLKKRGAANIKFVCLVAAPEGVKKLQEAHPDIDIF 180
 DI+QRQIF+VDPMLATGGSAILAVDSLKKRGAANIKFVCLVAAPEGVKKLQ+AHPDIDI+
 Sbjct: 121 DIDQRQIFVVDPMPLATGGSAILAVDSLKKRGAANIKFVCLVAAPEGVKKLQDAHPDIDIY 180
 Query: 181 TAALDDHLNEHGYIVPGLGDAGDRLFGTK 209
 TA+LD+ LNE+GYIVPGLGDAGDRLFGTK
 Sbjct: 181 TASLDEKLNENGYIVPGLGDAGDRLFGTK 209

An alignment of the GAS and GBS proteins is shown below.

Identities = 190/209 (90%), Positives = 201/209 (95%)

Query: 1 MGKQVISHPLIQHKLSILRRTTTSTKDFRELVDEIAMLMGYEVSRLPLEDVEIQTPVA 60
 MGK QVISHPLIQHKLSILRR TTSTKDFRELV+EIAMLMGYEVSRLPLEDV+IQTPV+
 Sbjct: 1 MGKQVISHPLIQHKLSILRRQTTSTKDFRELVNEIAMLMGYEVSRLPLEDVDIQTTPVS 60
 Query: 61 TTVQKQLAGKKLAIVPILRAGIGMVDGFLSLVPAAKVGHIGMYRDEETFQFVEYLVKLPE 120
 TVQKQLAGKKLAIVPILRAGIGMVDG LSLVPAAKVGHIGMYR+EET +PVEYLVKLPE
 Sbjct: 61 KTVQKQLAGKKLAIVPILRAGIGMVDGLLSLVPAAKVGHIGMYRNEETLEPVEYLVKLPE 120
 Query: 121 DIDQRQIFVVDPMPLATGGSAILAVDSLKKRGAASIKFVCLVAAPEGVAALQEAHPDVIDIY 180
 DI+QRQIF+VDPMLATGGSAILAVDSLKKRGAA+IKFVCLVAAPEGV LQEAHPD+DI+
 Sbjct: 121 DINQRQIFLVDPMPLATGGSAILAVDSLKKRGAANIKFVCLVAAPEGVKKLQEAHPDIDIF 180
 Query: 181 TAALDEKLNENGYIVPGLGDAGDRLFGTK 209
 TAALD+ LNEHGYIVPGLGDAGDRLFGTK
 Sbjct: 181 TAALDDHLNEHGYIVPGLGDAGDRLFGTK 209

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2011

A DNA sequence (GBSx2122) was identified in *S.agalactiae* <SEQ ID 6219> which encodes the amino acid sequence <SEQ ID 6220>. This protein is predicted to be hemolysin (patB). Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.29 Transmembrane 88 - 104 (86 - 106)
 ----- Final Results -----

-2263-

bacterial membrane --- Certainty=0.2317(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15133 GB:Z99120 aminotransferase [Bacillus subtilis]
 Identities = 130/381 (34%), Positives = 221/381 (57%), Gaps = 4/381 (1%)

10 Query: 5 DFTSLPERFSSNTIKWKAVQK---DQEILPLWIADMFPIPFEMSEAIEDFSHQMVFGYD 61
 +F ER + ++KW + + LP+W+ADMDF ++EA+++ +FGY
 Sbjct: 2 NFDKREERLGTQSVKWDKTGELFGVTDALPMWVADMDFRAPEAITEALKERLDHGIFGYT 61

15 Query: 62 SPKDSLYQAISNWEVQEHGYQFDKKSLLLIDGVVPAISVAIQAFTEKGDVAVLINTPVYPP 121
 +P A+ W HG++ + +S+ GVV A+S+A+QAFT+ GD V++ PVY P
 Sbjct: 62 TPDQKTKDAVCGWMQNRHGKVNPEISITFSPGVVTALSMVQAFTEPGDQVVVQPPVYTP 121

20 Query: 122 FARTIKYNNRHLVSNSSLNNNOYFEIDFKOLEKDIIENNVKLYIFCSPHNPGRVWTKGE 181
 F ++ N RH++ N LL + + IDF+ LE + + +V L+I C+PHNP GR W++ +
 Sbjct: 122 FYHMVEKNGRHHILHNPILLEKDGAYAIDFEDLETKLSDPSVTILFILCNPHNPSGRSWSRED 181

25 Query: 182 IQKIGDICKRYNVILVSDEIHQDLVLFNDNVHHSFNTVDSSFKELSVILSSATKTFNIAGT 241
 + K+G++C + V +VSDEIH DL+L+ + H F ++ F ++SV ++ +KTFNIA
 Sbjct: 182 LLKLGECLCEHGVTVVSDEIHSIDLMLYGHKHTPFASLSDDFADISVTCAPSKTFNIAGL 241

30 Query: 242 KNSFAIIEKLRSDFKRQIANNQOEISSLGLLATEVAFTKEQWLKALKMELEGSIEY 301
 + S II + R+ F N +++ + A E A++K WL L +E ++
 Sbjct: 242 QASAIIPDRLKRAKFSASLQRNGLGGLNAFAVTAIEAAYSKGGPWLELITYIEKNMNE 301

35 Query: 302 LYEQL-TQKTHIKVMKPEGTYLVWLDIFSAYNLTHLEIQEKLRYDAKLILNDGLTFGKEGK 360
 L T+ +K+MKP+ +YL+WLDIFSAY L+ E+Q+++ K+IL G +G G+
 Sbjct: 302 AEAFLSTELFKVKMMKPDASYLIWLDIFSAYGLSDAELQQRMLKKGKVVILEPGTKYGPGE 361

Query: 361 KHARINVAAPRSVIEEAVLRL 381
 R+N + +++ + R+
 Sbjct: 362 GFMRLNAGCSLATLQDGLRRI 382

There is also homology to SEQ ID 1006.

SEQ ID 6220 (GBS392) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 2; MW 46.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 83 (lane 5; MW 71kDa).

GBS392-GST was purified as shown in Figure 217, lane 4.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2012

45 A DNA sequence (GBSx2123) was identified in *S.agalactiae* <SEQ ID 6221> which encodes the amino acid sequence <SEQ ID 6222>. This protein is predicted to be rRNA methylase, SpoU family (cspR). Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1436(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55

The protein has homology with the following sequences in the GENPEPT database.

-2264-

>GP:AAB02738 GB:U58864 CspR [Bacillus subtilis]
Identities = 84/155 (54%), Positives = 120/155 (77%), Gaps = 3/155 (1%)

Query: 19 HIVLFEPQIPANTGNIARTCAATNAPLHIIRPMGFPIDDKMKRAGLDYWDKLDVSYFYD 78
H+VL++P+IPANTGNIARTCAATN LH+IRP+GF DDK +KRAGLDYW+ ++V ++D
Sbjct: 4 HVVLYQPEIPANTGNIARTCAATNTTLHLIRPLGFSTDDKMLKRAGLDYWEFVNVDYHDS 63

Query: 79 LEE-FMLSCRGKVLHLSKFADKVYSDENYND-DQDHYFMFGREDKGLPETFMREHAEKAL 136
LEE F +GK I+KF + ++ +Y D D+D++F+FGRE GLP+ ++ + ++ L
Sbjct: 64 LEELEFAYKKGKFFFITKFGQQPHTSFYDYLDEYFFVFGRETSGLPKDLIQNNMDRCL 123

Query: 137 RIPMNDEHVRSLNVSNTVCMIVYEALRQQSFNLE 171
R+PM EHVRSLN+SNT ++VYEALRQQ++ +L+
Sbjct: 124 RLPMT-EHVRSLNLSNTAAILVYEALRQQNYRDLK 157

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6223> which encodes the amino acid sequence <SEQ ID 6224>. Analysis of this protein sequence reveals the following:

Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2236(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 135/182 (74%), Positives = 150/182 (82%)

Query: 1 MNIETLTQKNHRSDSGRNHIVLFEPQIPANTGNIARTCAATNAPLHIIRPMGFPIDDKM 60
M + L KN + RNHIVLF+PQIP NTGNIARTCAATNAPLHII+PMGFPIDDKM
Sbjct: 13 MTTKELINKNDKVKKARNHIVLFQIPQNTGNIARTCAATNAPLHIKPMGFPIDDKM 72

Query: 61 KRAGLDYWDKLDVSYFYDGLLEEFMLSCRGKVLHLSKFADKVYSDENYNDQDHYFMFGRED 120
KRAGLDYWDKL++ FYD LE+F+ C G++HLISKFA YS Y D HYF+FGRED
Sbjct: 73 KRAGLDYWDKLELHFYDHLQFQINQCHGQLHLISKFAVNYSQATYADGDSHYFLFGRED 132

Query: 121 KGLPETFMREHAEKALRIPMNDEHVRSLNVSNTVCMIVYEALRQQSFNLELSHTYENDK 180
GLPE FMREHAEKALRIPMNDEHVRSLNVSNTVCM++YEALRQQ F LEL HTYE+DK
Sbjct: 133 TGLPEDFMREHAEKALRIPMNDEHVRSLNVSNTVCMVIYEALRQQGFQGLELKHTEYHDK 192

Query: 181 LK 182
LK
Sbjct: 193 LK 194

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2013

A DNA sequence (GBSx2124) was identified in *S.agalactiae* <SEQ ID 6225> which encodes the amino acid sequence <SEQ ID 6226>. Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -6.79 Transmembrane 82 - 98 (69 - 100)
INTEGRAL Likelihood = -6.48 Transmembrane 27 - 43 (24 - 47)
INTEGRAL Likelihood = -5.52 Transmembrane 132 - 148 (126 - 151)
INTEGRAL Likelihood = -5.10 Transmembrane 162 - 178 (161 - 185)

----- Final Results -----
bacterial membrane --- Certainty=0.3718(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-2265-

A related GBS nucleic acid sequence <SEQ ID 9411> which encodes amino acid sequence <SEQ ID 9412> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:CAB13143 GB:Z99110 similar to amino acid permease [Bacillus subtilis]
    Identities = 46/143 (32%), Positives = 81/143 (56%), Gaps = 1/143 (0%)

    Query: 3  FAYDGTIFVNIAPVKNPKKNLPLAFVIGPALILLSYLAFFYGLTQILGASFIMTTGND 62
              FAYDGW + + E+KNP+K LP A G ++ Y+ + L IL A+ I+T G +
10  Sbjct: 203 FAYDGWILLAAALGGEMKNPEKLLPRAMTGGLLIVTAIYIFINFALLHILSANEIVTLGEN 262

    Query: 63 AINYAANIIFGPSVGRLLSFIVILSVLGVANGLLLGTMRLPQAFARGWIK-SERMANIN 121
              A + AA ++FG G+L+S +I+S+ G NG +L R+ A AER + +E+++++
15  Sbjct: 263 ATSTAATMLFGSIGGKLISVGIIVSIFGCLNGKVLSPFRVSFAMAERKQLPFAEKLSHVH 322

    Query: 122 LKYQMSLPASLTVTAVAIFWLFV 144
              ++ A A+A+ + +
20  Sbjct: 323 PSFRTPWIAISFQIALALIMMLI 345

```

20 There is also homology to SEQ ID 3114.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2014

25 A DNA sequence (GBSx2125) was identified in *S. galactiae* <SEQ ID 6227> which encodes the amino acid sequence <SEQ ID 6228>. Analysis of this protein sequence reveals the following:

```

    Possible site: 20
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
30      bacterial cytoplasm --- Certainty=0.1849(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

35 A related GBS nucleic acid sequence <SEQ ID 9439> which encodes amino acid sequence <SEQ ID 9440> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:AAD23454 GB:AF117741 cochaperonin GroES [Streptococcus pneumoniae]
    Identities = 31/52 (59%), Positives = 42/52 (80%)

40  Query: 2  GDGIRTLTGELVAPSVABGDTVLVENGAGLEVKGNEKVTVVRESDIVAVVK 53
              G G+RTL G+LVAPSV GD VLVE AGL+VKDG+EK +V E++I+A+++
    Sbjct: 42 GQGVRTLNGDLVAPSVKTGDRVLVEAHAGLDVKGDEKYIIVGEANILAIIE 93

```

45 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6229> which encodes the amino acid sequence <SEQ ID 6230>. Analysis of this protein sequence reveals the following:

```

    Possible site: 50
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
50      bacterial cytoplasm --- Certainty=0.3290(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-2266-

An alignment of the GAS and GBS proteins is shown below.

Identities = 29/49 (59%), Positives = 39/49 (79%)

5 Query: 4 GIRTLTGELVAPSVAEAGDTVLVENGAGLEVKGNEKVTVVRESDIVAVV 52
G+RT+TG+ V PSV+ G VLVENG LEV +EKV+++RESDI+A+V
Sbjct: 60 GVRTITGDSVLPSSVSGQEVLVENGHDLEVTVDDEKVSIIRESDIIV 108

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 2015

A DNA sequence (GBSx2126) was identified in *S.agalactiae* <SEQ ID 6231> which encodes the amino acid sequence <SEQ ID 6232>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

15

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1272(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD23455 GB:AF117741 chaperonin GroEL [Streptococcus pneumoniae]

Identities = 472/539 (87%), Positives = 513/539 (94%), Gaps = 1/539 (0%)

25

Query: 1 MAKDIKFSADARSAMVRGVDILADTVKVTLGPKGRNVVLEKAFGSPLITNDGVTIAKEIE 60
M+K+IKFS+DARSAMVRGVDILADTVKVTLGPK RNVVLEK+FGSPLITNDGVTIAKEIE
Sbjct: 1 MSKEIKFSSDARSAMVRGVDILADTVKVTLGPKDRNVVLEKSFGSPLITNDGVTIAKEIE 60

30

Query: 61 LEDHFENMGAKLVSEVASKTNDIAGDGTATVLTQAIIVREGKKNVTAGANPIGIRRGIE 120
LEDHFENMGAKLVSE+ASKTNDIAGDGTATVLTQAIIVREG+KNVTAGANPIGIRRGIE
Sbjct: 61 LEDHFENMGAKLVSEIASKTNDIAGDGTATVLTQAIIVREGIKNVTAGANPIGIRRGIE 120

35

Query: 121 TAVSAAVEELKEIAQPVSGKEAIAQVAAVSSRSEKVGGEYISEAMERVGNMGVITIEESRG 180
TAV+AAVE LK A PV+ KEAI+QVAAVSSRSEKVGGEYISEAME+VG DGVITIEESRG
Sbjct: 121 TAVAAAVEALKNNAIPVANKEAISQVAAVSSRSEKVGGEYISEAMEKVGKDGVTIEESRG 180

40

Query: 181 METELEVVVEGMQFDRGYLSQYVMVDNEKMVSELENPYILITDKKISNIQEIPLPLEEVLK 240
METELEVVVEGMQFDRGYLSQYVMVD+EKMV++LENPYILITDKKISNIQEIPLPLE +L+
Sbjct: 181 METELEVVVEGMQFDRGYLSQYVMVDSEKMVADLENPYILITDKKISNIQEIPLPLESILQ 240

45

Query: 241 TNRPLLIADDDVDGEALPTLVLNKIRGTFNVVAVKAPGFGDRRKAMLEDIAILTGTVVT 300
+NRPLLIADDDVDGEALPTLVLNKIRGTFNVVAVKAPGFGDRRKAMLEDIAILTGTV+T
Sbjct: 241 SNRPLLIADDDVDGEALPTLVLNKIRGTFNVVAVKAPGFGDRRKAMLEDIAILTGTVIT 300

50

Query: 301 EDLGLDLKDATMQVLGQSAKVTVDKDSTVIVEGAGDSSAIANRVAIKSQMEATTSDFDR 360
EDLGL+LKDAT++ LGQ+A+VTVDKDSTVIVEGAG+ AI++RVA+IKSQ+E TTS+FDR
Sbjct: 301 EDLGLDLKDATIEALGQAARVTVDKDSTVIVEGAGNPEAISHRVAVIKSQIETTTSEFDR 360

55

Query: 361 EKLQERLAKLAGGVAVIKVGAATETELKEMKLRIEDALNATRAAVEEGIVSGGGTALNVV 420
EKLQERLAKL+GGVAVIKVGAATETELKEMKLRIEDALNATRAAVEEGIV+GGGTAL NV
Sbjct: 361 EKLQERLAKLSGGVAVIKVGAATETELKEMKLRIEDALNATRAAVEEGIVAGGGTALANV 420

Query: 421 IEKVAALKLNGDEETGRNIVLRALEEPVRQIAYNAGYEGSVIIERLKQSEIGTFNAANG 480
I A L+L GDE TGRNIVLRALEEPVRQIA+NAG+EGS++I+RLK +E+G GFNA G
Sbjct: 421 IPAEATLELTGDEATGRNIVLRALEEPVRQIAHNAGYEGSIVIDRLKNAELGIGFNAATG 480

60

Query: 481 EWVDMVTTGIIDPVKVSALQNAASVASLILTEAVVANKPEPEAPTAPAMDPSMMGG 539
EWV+M+ GIIDPVKV+RSALQNAASVASLILTEAVVANKPEP AP APAMDPSMMGG
Sbjct: 481 EWNMIDQGIIDPVKVSALQNAASVASLILTEAVVANKPEPVAP-APAMDPSMMGG 538

-2267-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6233> which encodes the amino acid sequence <SEQ ID 6234>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1070(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 491/543 (90%), Positives = 515/543 (94%), Gaps = 3/543 (0%)

```

Query: 1  MAKDIKFSADARSAMVRGVDILADTVKVTLGPKGRNVVLEKAFGSPLITNDGVTIAKEIE 60
Sbjct: 3  MAKDIKFSADARAAMVRGVDMLADTVKVTLGPKGRNVVLEKAFGSPLITNDGVTIAKEIE 62

Query: 61  LEDHFENMGAKLVSEVASKTNDIAGDGTTTATVLTQAIIVREGLKNVTAGANPIGIRRGIE 120
Sbjct: 63  LEDHFENMGAKLVSEVASKTNDIAGDGTTTATVLTQAIIVREGLKNVTAGANPIGIRRGIE 122

Query: 121 TAVSAAVEELKEIAQPVSGKEAIAQVAAVSSRSEKVG EYISEAMERVGN DGVITIEESRG 180
Sbjct: 123 TATATAVEALKAIAQPVSGKEAIAQVAAVSSRSEKVG EYISEAMERVGN DGVITIEESRG 182

Query: 181 METELEVEGGMQFDRGYLSQYMTDNEKVMSELENPFYILITDKKISNIQ EILPLLEEVLK 240
Sbjct: 183 METELEVEGGMQFDRGYLSQYMTDNEKVM++LENP+ILITDKK+SNIQ+ILPLLEEVLK 242

Query: 241 TNRPLLI IADDVDGEALPTLVLNKIRGTFNVVAVKAPGFGDRRKAMLEDIA IILTGGTVVT 300
Sbjct: 243 TNRPLLI IADDVDGEALPTLVLNKIRGTFNVVAVKAPGFGDRRKAMLEDIA IILTGGTVIT 302

Query: 301 EDLGLDLKD ATMQLGQSAKVTVDK DSTVIVEGAGDSSAIANRVAI IKSQMEATTSDFDR 360
Sbjct: 303 EDLGLLEKD ATM TALGQAAKITVDK DSTVIVEGSGSSRAIANRIAL IKSQLETTTSDFDR 362

Query: 361 EKLQERLAKLAGGVAVIKVGAATETELKEMKLRIEDALNATRAAVEEGIVSGGGTALVNV 420
Sbjct: 363 EKLQERLAKLAGGVAVIKVGAPTETALKEMKLRIEDALNATRAAVEEGIVAGGGTALITV 422

Query: 421 IEKVAALKLNGDEETGRNIVLRAL EEPVRQIAYNAGYEGSVI IERLKQSEIGTGFNAANG 480
Sbjct: 423 IEKVAALELEGDDATGRNIVLRAL EEPVRQIALNAGYEGSVVIDKLKNSPACTGFNAATC 482

Query: 481 EWVDMVT TGIIDPVKVTRSAIQNAASVASLIL TTEAVVANKPEP--EAPTAPA-MDP SMM 537
Sbjct: 483 EWVDMIKTGIIDPVKVTRSAIQNAASVASLIL TTEAVVANKPEPATPAPAMPAGMDPGMM 542

Query: 538 GGF 540
Sbjct: 543 GGF 545

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2016

A DNA sequence (GBSx2127) was identified in *S.agalactiae* <SEQ ID 6235> which encodes the amino acid sequence <SEQ ID 6236>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

-2268-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3216(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10247> which encodes amino acid sequence <SEQ ID 10248> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:BAB06113 GB:AP001515 transcriptional regulator (GntR family)
 [Bacillus halodurans]
 Identities = 50/171 (29%), Positives = 86/171 (50%), Gaps = 17/171 (9%)

 Query: 21 HVQVYNKIFNMIQDGTYSQGMQLPSEPELAGQLNVSRTLRLKSLALLQEDHLVKNIRGKG 80
 ++QV +K+ + ++ G Y G +LPSE EL+ QL VSRATLR++L LL+E+ +V G G
 15 Sbjct: 10 YLQVIDKLKHDMEAGVVEEGEKLPSFELSKQLGVSRTLRLALRLLEEEGVVRRHGVG 69

 Query: 81 NFIRESSNLSSENGYENRQHPIKTCITKITEVELE-----FRVEVPAAEITASLKQ 132
 F+ ++ L G E +T I ++E +++E +
 20 Sbjct: 70 TFW--HTKPLFSAGIEELY-----SVTDMIRHADMEPGTIFLSSYQIEATDDDKRRFQTD 122

 Query: 133 ETPVVVIADRWYHTDDGPLAYTSLFPIELISDAEISLHDTKQLLNFIIEEG 183
 +++ +R D P+ Y L +P ELI + S+H+ +L+ +E G
 25 Sbjct: 123 NLDQLMMIERVRTADGVPIVCLDKLPAELI--GQHSVHEINSILDHLESG 171

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6237> which encodes the amino acid sequence <SEQ ID 6238>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.2297(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 154/244 (63%), Positives = 189/244 (77%)

 Query: 7 MPKNELNKLNKLNKLVQVYNKIFNMIQDGTYSQGMQLPSEPELAGQLNVSRTLRLKSLAL 66
 M N+L KL KLKHVQVYN IF +IQDGTYSQGMQLPSEPELA QLNVSRTLRLKSLAL
 40 Sbjct: 1 MSTNDLTKKLKLVQVYNTIFQLIQDGTYSQGMQLPSEPELAGQLNVSRTLRLKSLAL 60

 Query: 67 LQEDHLVKNIRGKGNFIRESSNLSSENGYENRQHPIKTCITKITEVELEFRVEVPAAEI 126
 LQEDHL+KNIRGKGNFI + G+E QHPI L+S IT+VELE+R+EVP AI
 45 Sbjct: 61 LQEDHLIKNIRGKGNFILKTPETKYHQGFYELQHPIYASLSSDITKVELEYRIEVPTVAI 120

 Query: 127 TASLKQETPVVVIADRWYHTDDGPLAYTSLFPIELISDAEISLHDTKQLLNFIIEEGIYQ 186
 TASLKQETPVV+I DRWYH+ + +AY+LSFPIE+IS I+L+ + LL F+EE IY+
 50 Sbjct: 121 TASLKQETPVVIIIDRWYHSQNKALAYSLSFPIEIVISKYAINLNQEEPLTTFLEEKIYE 180

 Query: 187 EGISSHSQSHLGYATSGNFSATKYTLSDHCGFILIQETIFKQEKILMCNKHYVPPIEHFEL 246
 G +SHS + +GY +GN++ATKYTL++ FILLIET++ + IL+ KHYVP + F+L
 55 Sbjct: 181 SGKASHSCNQIGYTKTGNYTATKYTLSENSAFILLIETLYNGKDILVSTKHYVPADLFDL 240

 Query: 247 SITS 250
 + S
 Sbjct: 241 KVQS 244

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2269-

Example 2017

A DNA sequence (GBSx2128) was identified in *S.galactiae* <SEQ ID 6239> which encodes the amino acid sequence <SEQ ID 6240>. This protein is predicted to be purine nucleoside phosphorylase (udp-1). Analysis of this protein sequence reveals the following:

```

5   Possible site: 47
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3910 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:AAC65977 GB:AE001270 uridine phosphorylase (udp) [Treponema
      pallidum]
      Identities = 145/246 (58%), Positives = 171/246 (68%)

Query: 11 QYHLQIRPGDVGRYVIMPGDPKRCAKIAEHFDNAVLVADSREYVTTYGTLNGEKVSVTST 70
      +YH+ ++ D+G YVI+PGDP R KIA+HF + V +REYVTTYGTL VSV ST
20   Sbjct: 10 EYHIGLKASDIGHYVILEPGDPARSEKIAQHFSHPHKVGHNREYVTTYGTLCETPVSVMT 69

Query: 71 GIGGPSASIAM EELKLCGADTFIRVGTCCGIDLDVKG GDIVIATGAIRMEGTSKEYAPIE 130
      GIGGPS +I +EEL GA TFIRVGT GG+ D+ G +VIATGAIR EGTSKEYAP+E
25   Sbjct: 70 GIGGPSTAIGVEELIHLGAHTFIRVGTSGGMQPDILAGTVVIATGAIRFEGTSKEYAPVE 129

Query: 131 FPAVADLEV TNALVNAAK KLYTSHAGVVQCKDAFYGQHEPERMPVSYELLNKWEAWKRL 190
      FPAV D VT AL +AA+ + GVVQCKD FYGOH P MPV EL KW AW
30   Sbjct: 130 FPAVPDFTVTAALKHAAEDVQVRHALGVVQCKDNFYGQHSPTMPVHAELTQKWHAWIAC 189

Query: 191 GTKASEMESAA LFVAASHLGVRCGSDFLVGNQERNALGMDNPM AHDTEAAIQVAVEALR 250
      T ASEMESAA LFV S VR G+ LV+GNQ R A G+++ HDTE AI+VAVEA++
35   Sbjct: 190 NTLASEMESAA LFVLGSVRRVRTGAVLLVIGNQTRRAQGLEDIQVHDTENAIRVAVEAVK 249

Query: 251 TLIEND 256
      LI D
40   Sbjct: 250 LLITQD 255

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6241> which encodes the amino acid sequence <SEQ ID 6242>. Analysis of this protein sequence reveals the following:

```

40   Possible site: 47
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
45      bacterial cytoplasm --- Certainty=0.3910 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

50   Identities = 259/259 (100%), Positives = 259/259 (100%)

Query: 1 MQNYSGEVGLQYHLQIRPGDVGRYVIMPGDPKRCAKIAEHFDNAVLVADSREYVTTYGTL 60
      MQNYSGEVGLQYHLQIRPGDVGRYVIMPGDPKRCAKIAEHFDNAVLVADSREYVTTYGTL
Sbjct: 1 MQNYSGEVGLQYHLQIRPGDVGRYVIMPGDPKRCAKIAEHFDNAVLVADSREYVTTYGTL 60

55   Query: 61 NGEKVSVTSTGIGGPSASIAM EELKLCGADTFIRVGTCCGIDLDVKG GDIVIATGAIRME 120
      NGEKVSVTSTGIGGPSASIAM EELKLCGADTFIRVGTCCGIDLDVKG GDIVIATGAIRME
Sbjct: 61 NGEKVSVTSTGIGGPSASIAM EELKLCGADTFIRVGTCCGIDLDVKG GDIVIATGAIRME 120

60   Query: 121 GTSKEYAPIEFPAVADLEV TNALVNAAK KLYTSHAGVVQCKDAFYGQHEPERMPVSYEL 180
      GTSKEYAPIEFPAVADLEV TNALVNAAK KLYTSHAGVVQCKDAFYGQHEPERMPVSYEL
Sbjct: 121 GTSKEYAPIEFPAVADLEV TNALVNAAK KLYTSHAGVVQCKDAFYGQHEPERMPVSYEL 180

```

-2270-

Query: 181 LNKWEAWKRLGTKASEMESAAALFVAASHLGVRCSDFLVVGNQERNALGMDNPMAMDTEA 240
 LNKWEAWKRLGTKASEMESAAALFVAASHLGVRCSDFLVVGNQERNALGMDNPMAMDTEA
 Sbjct: 181 LNKWEAWKRLGTKASEMESAAALFVAASHLGVRCSDFLVVGNQERNALGMDNPMAMDTEA 240

Query: 241 AIQVAVEALRTLIENTDKSQ 259
 AIQVAVEALRTLIENTDKSQ
 Sbjct: 241 AIQVAVEALRTLIENTDKSQ 259

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2018

A DNA sequence (GBSx2129) was identified in *S.agalactiae* <SEQ ID 6243> which encodes the amino acid sequence <SEQ ID 6244>. This protein is predicted to be nucleoside transporter. Analysis of this protein sequence reveals the following:

Possible site: 25

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -9.45	Transmembrane	35 - 51 (30 - 57)
INTEGRAL	Likelihood = -9.29	Transmembrane	8 - 24 (1 - 28)
INTEGRAL	Likelihood = -8.07	Transmembrane	388 - 404 (379 - 404)
INTEGRAL	Likelihood = -7.27	Transmembrane	104 - 120 (100 - 127)
INTEGRAL	Likelihood = -6.58	Transmembrane	259 - 275 (255 - 284)
INTEGRAL	Likelihood = -4.35	Transmembrane	172 - 188 (171 - 190)
INTEGRAL	Likelihood = -3.50	Transmembrane	200 - 216 (199 - 221)
INTEGRAL	Likelihood = -2.18	Transmembrane	352 - 368 (352 - 371)

----- Final Results -----

bacterial membrane	---	Certainty=0.4779(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 10245> which encodes amino acid sequence <SEQ ID 10246> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05165 GB:AP001512 nucleoside transporter [Bacillus halodurans]
 Identities = 160/405 (39%), Positives = 256/405 (62%), Gaps = 8/405 (1%)

Query: 5 MQFIYSIIIGILLVLGIVYAIISFNKRSVLSLIGKALIVQFIIALILVRIPLGQQVSVSVS 64
 M ++ ++GI++V I +A S NR+++ I L +Q + A+I+++IP GQ ++ ++
 Sbjct: 1 MNILWGLLGIVVFLIAFAFSTNRRRAIKPRITLGLAIQLLFAIIVLKIPAGQALLESIT 60

Query: 65 TGVTRVINCGQAGLNFVFGSLADSGAKTGFIQTLGNIVFLSALVSLLYVVGILGFVV 124
 V +I+ G++FVFG + G+ GF+FAI L ++F SAL+S+LYY+GI+ FV+
 Sbjct: 61 NVVLNIISYANEGIDFVFGGFFEEGSGVGFVFAINVLSVVIFFSALISILYILGIMQFVI 120

Query: 125 KWIGKGVGKIMKSSEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGMGMS 184
 K IG + ++ +S+ ES A AN+F+GQT++P++V YL +MT SE+ V+ G+ S++
 Sbjct: 121 KIIGGALSLLGTSKAESMSAAAANIFVGQTEAPLVVVKPYLPKMTQSELFVMTGGLASVA 180

Query: 185 VSILGGYIALGIPMEYLLIASTMPVIGSILIAKILLPQTEPVQKI-DDIKMDNKGNNANV 243
 S+L GY LG+P++YLL AS M +++AK+++P+TE DD K+ + N+
 Sbjct: 181 GSVLIGYSLGLVPLQYLLAASFMAAPAGLIMAKMIMPETEKTDAEDDFKLAKDEESTNL 240

Query: 244 IDAIAEGASTGAQMAFSIGASLIAFVGLVSLINMMLSGLG-----IRLEQIFS YVFAP 296
 IDA A GASTG + +I A L+AFV L++LIN +L +G + LE I YVFAP
 Sbjct: 241 IDAAANGASTGLMLVLNIAAMLLAFVALIALINGILGWIGGLFGASQLSLELILGYVFAP 300

Query: 297 FGFLMGFDHKNILLEGNLLGSKLILNEFVSFQQLGDLIKSLDYRTALVATISLCGFANLS 356
 F++G L G+ +G KL++NEFV++ I++L + +V + +LCGFAN S

-2271-

Sbjct: 301 LAFVIGIPWAEALQAGSYIGQKLVVNEFVAYLSFAPEIENLSDKAVMVISFALCGFANFS 360

Query: 357 SLGICVSGIAVLCPEKRGTLARLVFRAMIGGIASMLSAFIVGIV 401

SLGI + G+ L P +R +ARL RA++ G S+LSA I G++

5 Sbjct: 361 SLGILLGGLGKLAPSRPDIARLGLRAILAGTLASLLSASIAGML 405

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6245> which encodes the amino acid sequence <SEQ ID 6246>. Analysis of this protein sequence reveals the following:

10 Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -9.45 Transmembrane 35 - 51 (30 - 57)
 INTEGRAL Likelihood = -9.29 Transmembrane 8 - 24 (1 - 28)
 INTEGRAL Likelihood = -8.07 Transmembrane 388 - 404 (379 - 404)
 INTEGRAL Likelihood = -7.27 Transmembrane 104 - 120 (100 - 127)
 15 INTEGRAL Likelihood = -6.58 Transmembrane 259 - 275 (255 - 284)
 INTEGRAL Likelihood = -4.35 Transmembrane 172 - 188 (171 - 190)
 INTEGRAL Likelihood = -3.50 Transmembrane 200 - 216 (199 - 221)
 INTEGRAL Likelihood = -2.18 Transmembrane 352 - 368 (352 - 371)
 20 ----- Final Results -----
 bacterial membrane --- Certainty=0.4779(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:

>GP:BAB05165 GB:AP001512 nucleoside transporter [Bacillus halodurans]

Identities = 160/405 (39%), Positives = 257/405 (62%), Gaps = 8/405 (1%)

30 Query: 5 MQFIYSIIIGILLVLGIVYAISFNKRSVSLSLIGKALIVQFIIALILVRIPLGQQIVSVVS 64
 M ++ ++GI++V I +A S NR+++ I L +Q + A+I+++IP GQ ++ ++
 Sbjct: 1 MNILWGLLGIVVFLIAFAFSTNRRRAIKPRTILGGLAIQLLFAIIVLKIPAGQALLESIT 60
 Query: 65 TGVTSVINCGQAGLNFVFGSLADSGAKTGFIATQTLGNIVFLSALVSLLYYVVGILGFVV 124
 V ++I+ G++FVFG + G+ GF+FAI L ++F SAL+S+LYY+GI+ FV+
 35 Sbjct: 61 NVVLNIISYANEGIDFVGGFFEEGSGVGVFPAINVLSVVIFFSALISILYYLQIMQFVI 120
 Query: 125 KWIGKGVGKIMKSSEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGMGMSMS 184
 K IG + ++ +S+ ES A AN+F+GQT++P++V YL +MT SE+ V+ G+ S++
 40 Sbjct: 121 KIIGGALSWLGTSKAESMSAAANIFVGGTEAPLVVVKPYLPKMTQSELFAMVTGGLASVA 180
 Query: 185 VSILGGYIALGIPMEYLLIASTMVPIGSILIAKILLPQTEPVQKI-DDIKMDNKGNNANV 243
 S+L GY LG+P++YLL AS M +++AK+++P+TE DD K+ + N+
 Sbjct: 181 GSVLIGYSLLGVPLQYLLAASFMAAPAGLIMAKMIMPETEKTTDAEDDFKLAKDEESTNL 240
 45 Query: 244 IDAIAEGASTGAQMAFSIGASLIAFVGLVSLINMMLSGLG-----IRLEQIFS YVFAP 296
 IDA A GASTG + +I A L+AFV L++LIN +L +G + LE I YVFAP
 Sbjct: 241 IDAAANGASTGLMLVLNIAAMLLAFVALIALINGILGWIGGLFGASQLSLELILGYVFAP 300
 Query: 297 FGFLMGFDHKNILLEGNLLGSKLILNEFVSFQQLGHLIKSLDYRTALVATISLCGFANLS 356
 F++G L G+ +G KL++NEFV++ I++L + +V + +LCGFAN S
 50 Sbjct: 301 LAFVIGIPWAEALQAGSYIGQKLVVNEFVAYLSFAPEIENLSDKAVMVISFALCGFANFS 360
 Query: 357 SLGICVSGIAVLCPEKRSTLARLVFRAMIGGIASMLSAFIVGIV 401
 SLGI + G+ L P +R +ARL RA++ G S+LSA I G++
 55 Sbjct: 361 SLGILLGGLGKLAPSRPDIARLGLRAILAGTLASLLSASIAGML 405

An alignment of the GAS and GBS proteins is shown below.

Identities = 399/404 (98%), Positives = 401/404 (98%)

60 Query: 1 MEVIMQFIYSIIIGILLVLGIVYAISFNKRSVSLSLIGKALIVQFIIALILVRIPLGQQVV 60
 +EVIMQFIYSIIIGILLVLGIVYAISFNKRSVSLSLIGKALIVQFIIALILVRIPLGQQ+V
 Sbjct: 1 LEVIMQFIYSIIIGILLVLGIVYAISFNKRSVSLSLIGKALIVQFIIALILVRIPLGQQIV 60
 Query: 61 SVVSTGVTKVINCGQAGLNFVFGSLADSGAKTGFIATQTLGNIVFLSALVSLLYYVVGIL 120

		SVVSTGVT VINCGQAGLNFVFGSLADSGAKTGFIFAIQTGLGNIVFLSALVSLLYYVGIL	
Sbjct:	61	SVVSTGVTSVINCGQAGLNFVFGSLADSGAKTGFIFAIQTGLGNIVFLSALVSLLYYVGIL	120
Query:	121	GFVVKWIGKGVGKIMKSSEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGM	180
		GFVVKWIGKGVGKIMKSSEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGM	
Sbjct:	121	GFVVKWIGKGVGKIMKSSEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGM	180
Query:	181	GSMSVSILGGYIALGIPMEYLLIASTMVPIGSILIAKILLPQTEPVQKIDDIKMDNKGNN	240
		GSMSVSILGGYIALGIPMEYLLIASTMVPIGSILIAKILLPQTEPVQKIDDIKMDNKGNN	
Sbjct:	181	GSMSVSILGGYIALGIPMEYLLIASTMVPIGSILIAKILLPQTEPVQKIDDIKMDNKGNN	240
Query:	241	ANVIDATAEGASTGAQMAFSIGASLIAFVGLVSLINMMSGLGIRLEQIFSYPAPFGFL	300
		ANVIDATAEGASTGAQMAFSIGASLIAFVGLVSLINMMSGLGIRLEQIFSYPAPFGFL	
Sbjct:	241	ANVIDATAEGASTGAQMAFSIGASLIAFVGLVSLINMMSGLGIRLEQIFSYPAPFGFL	300
Query:	301	MGFDHKNILLEGNLLGSKLIILNEFVSFQQLGDLIKSLDYRTALVATISLCGFANLSSLGI	360
		MGFDHKNILLEGNLLGSKLIILNEFVSFQQLGLIKSLDYRTALVATISLCGFANLSSLGI	
Sbjct:	301	MGFDHKNILLEGNLLGSKLIILNEFVSFQQLGHLIKSLDYRTALVATISLCGFANLSSLGI	360
Query:	361	CVSGIAVLCPEKRGTLARLVFRAMIGGIAVMSLSAFIVGIVTLF	404
		CVSGIAVLCPEKR TLARLVFRAMIGGIAVMSLSAFIVGIVTLF	
Sbjct:	361	CVSGIAVLCPEKRSTLARLVFRAMIGGIAVMSLSAFIVGIVTLF	404

Lipop: Possible site: -1 Crend: 1
McG: Discrim Score: 13.83
GvH: Signal Score (-7.5): -2.63
 Possible site: 25

```
>>> Seems to have an uncleavable N-term signal seq
ALOM program      count: 8 value: -9.45 threshold: 0.0
INTEGRAL          Likelihood = -9.45 Transmembrane 35 - 51 ( 30 - 57)
INTEGRAL          Likelihood = -9.29 Transmembrane 8 - 24 ( 1 - 28)
INTEGRAL          Likelihood = -8.07 Transmembrane 388 - 404 ( 379 - 404)
INTEGRAL          Likelihood = -7.27 Transmembrane 104 - 120 ( 100 - 127)
INTEGRAL          Likelihood = -6.58 Transmembrane 259 - 275 ( 255 - 284)
INTEGRAL          Likelihood = -4.35 Transmembrane 172 - 188 ( 171 - 190)
INTEGRAL          Likelihood = -3.50 Transmembrane 200 - 216 ( 199 - 221)
INTEGRAL          Likelihood = -2.18 Transmembrane 352 - 368 ( 352 - 371)
PERIPHERAL        Likelihood = 3.82 286
modified ALOM score: 2.39
```

*** Reasoning Step: 3

```

----- Final Results -----
      bacterial membrane --- Certainty=0.4779(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

ORF01622(313 - 1512 of 1812)
 GP|9656920|gb|AAF95495.1||AE004305(1 - 418 of 418) NupC family protein {Vibrio cholerae}
 %Match = 24.0
 %Identity = 39.5 %Similarity = 65.7
 Matches = 160 Mismatches = 134 Conservative Sub.s = 106

276 306 336 366 396 426 456 486
C*STPHTY*K*ITISEVLEVINQFIYSIIIGILLVLGIVYAI SFNRKSVSLSLIGKALIVQFTIALILVRIPLGQQVSVS
| : | : | : : | | : | | | : : | : | : | : | : : : | | : :
MSLFMSLIGMAVLLGLIAVLLSSNRKAINLRTVGGAFAIQFSLGAFILYVPWQGELLRG
10 20 30 40 50

VSTGVTKVINCGQAGLN⁵¹⁶FVFGSLADSG-----AKTGFIFAIQT⁵⁹¹LGNIVFLSALVSLLYYVGILGVFKWIGKGVKGMKS
| : || | : : || | : : : | : : : | : : : | : : :

516 546 591 621 651 681 711

-2273-

```

FSDAVSNVINYGNDGTSFLFGGLVSGKMFVFGGGGFIFAFRVLPTLIFFSALISVLYLGVMQWVIRILGGGLQKALGT
      70      80      90     100     110     120     130

741      771      801      831      861      891      921      951
5 SEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGMGMSVSILGGYIALGIPMEYLLIASTMVPIGSILIAK
  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
SRAESMSAAANIFVGQTEAPLVVRPFVPKMTQSELEFAVMCGGLASLAGGVLAGYASMGVKIEYLVAASEFMAAPGGLLFAK
      150     160     170     180     190     200     210

10 981      1011     1038     1068     1098     1128              1167
  ILLPQTEFPVQKIDDIKMDNKGNN-ANVIDATAEGASTGAQMAFSIGASLIAFVGLVSLINMMLSGLG-----IRLEQI
  ::|:|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
LMPPETEKPDNEDITLDGGDDKPANVIDAAAGGASAGLQALNVGAMLIAPFIGLIALINGMLGGIGGWFGMPCLKLEML
      230     240     250     260     270     280     290

15 1197     1227     1257     1287     1305     1332     1362     1392
  FSYVFAPFGFLMGFDHKNILLEGNLLGSKLILNEFVSFQQ----LGDLIKSLDYRTALVATISLCGFANLSSLGICVSG
  :  ::|:|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
LGWLFAPLAFILGVPWNEATVAGEFIGLKTVA NEFVAYSQFAPYLTEAAPVVLSEKTKAII SFALCGFANLSSIAILLGG
20      310     320     330     340     350     360     370

1422     1452     1482     1512     1542     1572     1602     1632
  IAVLCPEKRGTLARLVFRAMIGGIAVSMLSAFIVGIVTLF*KLTKERRIVTWK*KIF*KR*TILC*QQQHQGQSKQF*M
  :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
25 LGS LAPKRRGDIARMGVKAVIAGTLSNLMAATIAGFFLSF
      390     400     410

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

30 Example 2019

A DNA sequence (GBSx2130) was identified in *S. galactiae* <SEQ ID 6247> which encodes the amino acid sequence <SEQ ID 6248>. This protein is predicted to be deoxyribose-phosphate aldolase (deoC). Analysis of this protein sequence reveals the following:

```

35 Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2196(Affirmative) < succ>
40      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA81646 GB:Z27121 deoxyribose aldolase [Mycoplasma hominis]
Identities = 99/199 (49%), Positives = 140/199 (69%), Gaps = 1/199 (0%)
45
Query: 5 DILKTV DHTLLATTATWPEIQITILDDAMAYETASACIPASYVKKAAEYVSGK-LAICTVI 63
++ K +DHT L+ +AT +I ++ +A+ Y+ S CI SYVK A E + + +CTVI
Sbjct: 3 ELNKYIDHTNLSFSATS KDIDKLIQEAIKYDFKSVCIAPSYVKYAKEALKNSDVLVCTVI 62

50
Query: 64 GFPNGYSTTAAKVFECDQDAIKNGADEIDMVINLTDVKNQDFDTEEEIRQIKAAACQDHIL 123
GFP GY+ T+ KV+E + A+++GADEIDMVIN+ K+G ++ V EI+ IK AC L
Sbjct: 63 GFP LGYNATSVKVYETKIAVEHGADEIDMVINVGFRKDGQYEVVLNEIKAIKEACNGKTL 122

55
Query: 124 KVIVETCQLTKHEELIELCGVVTRSGADFIKTSTGFSTAGATFEDVEVMAKYVGEVGIKA 183
KVIVET LTK ELI++ +V +SGADFIKTSTGFS GA+FED++ M + G+ + IKA
Sbjct: 123 KVIVETALLTKAELIKITELVMQSGADFIKTSTGFSYRGASFEDIQTMKETCGDKLLIKA 182

Query: 184 AGGISSLEDAEKFIALGAS 202
+GGI +L DA++ I LGA+
60 Sbjct: 183 SGGIKNLADAQEMIRLGAN 201

```

-2274-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6249> which encodes the amino acid sequence <SEQ ID 6250>. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2196(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 211/223 (94%), Positives = 217/223 (96%)

Query: 1 MEVKDILKTV DHTLLATTATWPEIQ TILDDAMAYETASACIPASYVKKAAEYVSGKLAIC 60

+EVKDILKTV DHTLLATTATWPEIQ TILDDAMAYETASACIPASYVKKAAEYVSGKLAIC

Sbjct: 1 VEVKDILKTV DHTLLATTATWPEIQ TILDDAMAYETASACIPASYVKKAAEYVSGKLAIC 60

Query: 61 TVIGFPGYSTTAAKVFECDQDAIKNGADEIDMVINLTDVKNQDFDFTVEEEIRQIKAAQCQD 120

TVIGFPGYSTTAAKVFECDQDAI+NGADEIDMVINLTDVKNQDFDFTVEEEIRQIKAA CQD

Sbjct: 61 TVIGFPGYSTTAAKVFECDQDAIQNGADEIDMVINLTDVKNQDFDFTVEEEIRQIKAKCQD 120

Query: 121 HILKVIVETCQLTKEELIELCGVVTRSGADFIKTSTGFSSTAGATFEDVEVMAKYVGEVVK 180

HILKVIVETCQLTKEELIELCGVVTRSGADFIKTSTGFSSTAGATFEDVEVMAKYVGEVVK

Sbjct: 121 HILKVIVETCQLTKEELIELCGVVTRSGADFIKTSTGFSSTAGATFEDVEVMAKYVGEVVK 180

Query: 181 IKAAGGISSLEDAEKFIALGASRLGTSRIIKIVKNQKVEEGTY 223

IKAAGGISSLEDA+ FIALGASRLGTSRIIKIVKN+ + +Y

Sbjct: 181 IKAAGGISSLEDAKTFIALGASRLGTSRIIKIVKNEATKTDSY 223

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2020

A DNA sequence (GBSx2131) was identified in *S.galactiae* <SEQ ID 6251> which encodes the amino acid sequence <SEQ ID 6252>. This protein is predicted to be phosphopentomutase (deoB). Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0546(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45496 GB:U80410 phosphopentomutase [Lactococcus lactis subsp. cremoris]

Identities = 275/408 (67%), Positives = 325/408 (79%), Gaps = 7/408 (1%)

Query: 3 QFDRIHLVVLD SVGIGAAPDANDFVNAGVP-----DGASDTLGHISKTVGLAVPNMAKI 56

+F RIHLVV+DSVGIGAAPDA+ F N V D SDT+GHIS+ GL VPN+ K+

Sbjct: 4 KFGRIHLVMD SVGIGAAPDADKFFNHVDVETHEAINDVKSDTIGHISEIRGLDVPNLQKL 63

Query: 57 GLGNIPRPQALKTVPAEENPSGYATKLQEVSLGKDTMTGHWEIMGLNITEPFDFWNGFP 116

G GNIPR LKT+PA + P+ Y TKL+E+S GKDTMTGHWEIMGLNI PF T+ G+P

Sbjct: 64 GWGNIPRESPLKTI PAQAQKPAAYVTKLEEISKGKDTMTGHWEIMGLNIQTFFPTYPEGY 123

Query: 117 EDIITKIEDFSGRKVIREANKPYSGTAVIDDGPRQMETGELIITYTSADPVLQIAAHEDI 176

ED++ KIE+FSGRK+IREANKPYSGTAVI+DFGPRQ+ETGELIITYTSADPVLQIAAHED+

-2275-

Sbjct: 124 EDLLEKIEBFSGRKIIREANKPYSGTAVIEDFGPRQLETGELIITYTSADPVLQIAAHEDV 183

Query: 177 IPLEELYRICEYARSITMERPAL-LGRIIARPYVGE PGNFTRTANRHDYAVSPFEDTVLN 235
I EELY+ICEY RSIT+E ++ GRIIARPYVGE GNF RT R DYA+SPF +TVL

5 Sbjct: 184 ISREELYKICEYVRSITLEGSGIMIGRIIARPYVGEAGNFERTDGRDYALSPFAETVLE 243

Query: 236 KLDQAGIDTYAVGKINDIFNGSGINHDMGHNKSNSHGIDTLIKTMGLSEFEKGFSFTNLV 295
KL +AGIDTY+VGKI+DIFN G+ +DMGHN ++ G+D L+K M +EF +GFSFTNLV

10 Sbjct: 244 KLYKAGIDTYSVGKISDIFNTVGKIDMGHNHNDMDGVDRLKAMTKTEFTGFSFTNLV 303

Query: 296 DFDALYGHRRDPHGYRDCLEHFDERLPEIISAMRDKDLLITADHGNDPTYAGTDHTREY 355
DFDA YGHRRD GY + +FD RLPEII AM++ DLL+ITADHGNDP+Y GTDHTREY

Sbjct: 304 DFDALYGHRRDVEGYGAIEDFDGRLPEIIDAMKEDDLLMITADHGNDPSYVGTHTREY 363

15 Query: 356 IPLLAYSPSFTGNGLIPVGHFADISATVADNFGVDTAMIGESFLQDLV 403
IPL+ +S SF ++PVGHFADISAT+A+NF V A GESFL LV

Sbjct: 364 IPLVIFSKSFKEPKVLPVGHFADISATIAENFVKKQOTGESFLDALV 411

There is also homology to SEQ ID 2740:

20 Identities = 348/402 (86%), Positives = 374/402 (92%)

Query: 1 MSQFDRIHLLVVLDSVGIGAAPDANDFVNAGVPDGASDTLGHISKT VGLAVPNMAKIGLGN 60
MS+F+RIHLVVLDSVGIGAAPDA+ F NAGV D SDTLGHIS+ GL+VPNMAKIGLGN

25 Sbjct: 1 MSKFNRIHLVVLDSVGIGAAPDADKFFNAGVADTSDTLGHISEAGLSVPNMAKIGLGN 60

Query: 61 IPRPQALKTPAEENPSGYATKLQEVSLGKDTMIGHWEIMGLNITEPFDTFWNGFPEDII 120
I RP LKTVP E+NP+GY TKL+EVSLGKDTMIGHWEIMGLNITEPFDTFWNGFPE+I+

Sbjct: 61 ISRPIPLKTVPTEDNPTGYVTKLEEVSLGKDTMIGHWEIMGLNITEPFDTFWNGFPPEIL 120

30 Query: 121 TKIEDFSGRKVIREANKPYSGTAVIDDGPRQMETGELIITYTSADPVLQIAAHEDIIPLE 180
TKIE+FSGRK+IREANKPYSGTAVIDDGPRQMETGELI+YTSADPVLQIAAHEDIIP+E

Sbjct: 121 TKIEEFSGRKIIREANKPYSGTAVIDDGPRQMETGELIVYTSADPVLQIAAHEDIIPVE 180

35 Query: 181 ELYRICEYARSITMERPALLGRIIARPYVGE PGNFTRTANRHDYAVSPFEDTVLNKLDQA 240
ELY+ICEYARSIT+ERPALLGRIIARPYVG+PGNFTRTANRHDYAVSPF+DTVLNKL A

Sbjct: 181 ELYKICEYARSITLERPALLGRIIARPYVGDPGNFTRTANRHDYAVSPFQDTVLNKLADA 240

Query: 241 GIDTYAVGKINDIFNGSGINHDMGHNKSNSHGIDTLIKTMGLSEFEKGFSFTNLVDFDAL 300
G+ TYAVGKINDIFNGSGI +DMGHNKSNSHGIDTLIKT+ L EF KGFSFTNLVDFDA

40 Sbjct: 241 GVPTYAVGKINDIFNGSGITNDMGHNKSNSHGIDTLIKTLQLPEFTKGFSFTNLVDFDAN 300

Query: 301 YGHRRDPHGYRDCLEHFDERLPEIISAMRDKDLLITADHGNDPTYAGTDHTREYIPLLA 360
+GHRRDP GYRDCLEHFD RLPEII+ M++ DLLITADHGNDPTYAGTDHTREYIPLLA

45 Sbjct: 301 FGHRDPFGYRDCLEHFDNRLEPIIANMKEDDLLITADHGNDPTYAGTDHTREYIPLLA 360

Query: 361 YSPSFTGNGLIPVGHFADISATVADNFGVDTAMIGESFLQDL 402
YS SFTGNGLIP GHFADISATVA+NFGVDTAMIGESFL L

Sbjct: 361 YSVSFTGNGLIPQGHFADISATVAENFGVDTAMIGESFLSHL 402

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2021

A DNA sequence (GBSx2132) was identified in *S.galactiae* <SEQ ID 6253> which encodes the amino acid sequence <SEQ ID 6254>. Analysis of this protein sequence reveals the following:

55 Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood =-12.05 Transmembrane 9 - 25 (4 - 35)

60 ----- Final Results -----
bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-2276-

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6255> which encodes the amino acid sequence <SEQ ID 6256>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 56
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -5.57    Transmembrane    41 - 57 ( 38 - 60)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.3230(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related sequence was also identified in GAS <SEQ ID 9143> which encodes the amino acid sequence <SEQ ID 9144>. Analysis of this protein sequence reveals the following:

```

15  Possible cleavage site: 49
   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -5.57    Transmembrane    13 - 29 ( 10 - 32)

20  ----- Final Results -----
      bacterial membrane --- Certainty= 0.323(Affirmative) < succ>
      bacterial outside --- Certainty= 0.000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

```

25 An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 276/544 (50%), Positives = 368/544 (66%), Gaps = 5/544 (0%)

Query: 5   FKKKVVKVCLVIFGIVLVSLLSLGGFFYFSKGQVLSRFVAARSRTSGQAFDNIKEYMVWSD 64
          F K +K +I L L G FY+SK ++ ++ ARS SG F+NIK Y+VW D
30  Sbjct: 33 FHHKKLKQITIIAATSLFLFLIGGAFYYSKNHCINAYLKARSAQSGPVFENIKAYLVWDD 92

Query: 65  TGESITNDEANYANFEPLSKSEARKLGQEIKEGNKNSMYLKRVGSRGIFPDYRIANKP 124
          T E ITNDEA Y F S+ E R+ Q++K +++ ++ +K VG R IFPDYRIA KP
35  Sbjct: 93 TNEQITNDEAMYTKFRYSQKELRQKKQDLKAASQDSAVQVKS VGRRFWIFPDYRIA IKP 152

Query: 125  MSLLTKTNVNPVKLDVLLNQKKVATSNSDHFSVTVERLPRTHYTASLEGTS DGKEIKLKKDY 184
          M LT+KTNVP+ DVLLNQKKVA S+S+ FSV ++RLP YTAS+ G +G+ IK+ K Y
40  Sbjct: 153 MDLTIKTNVPQADVLLNQKIVAVSDSEQFSVKLDRLPTAETASIRGIKHNGRNIKVNKSY 212

Query: 185  DGNQITIDLSVAFKSFTVTNSLMDGNLYFGDNRIAKLDGSHSVENYPVTDGSKAYIKKV 244
          DG N +DLSV+F++F VTSN G+LYF DN I LKDG VE+YPVT+ ++AY+K
45  Sbjct: 213 DGDNPVLDLSVSFRTFLVTSNAKQGDLYFDDNHIGTLKDGQLQVEDYPVTENAQAYMKTT 272

Query: 245  FNDGEITSHKQKLISIANQTIKLDVDGLNEKEAGQKLITAFNQLILYVSTGQDPQTLG 304
          F.DGE+ S K L + + T+++ V LL E +AG+ L++AF+QL+ Y+STGQD L
50  Sbjct: 273 FPDGELRSQKYALADVEEGATLEILVTDLLEEDKAGELLVSAFDQLMHYLS TGQDSSNLR 332

Query: 305  TVFEKGAENDFYKGLKESIKAKFVTDNRKASHFTIPNIVLNKMTQVGKESYQVNFAADYD 364
          +VFE G+ N FY+GLKESIKAKF TD RKAS IP+I+L MTQVGK +Y ++F A Y+
55  Sbjct: 333 SVFEAGSSNAFYRGLKESIKAKFQTDTRKASRLNIPSILLTMTQVGKTTYVLDFTATYE 392

Query: 365  FNYDKSTDPDKKTYGHIIQNLTGNFIMKKSNGSYLISNDGKKDITVAKETNKVKADPVSI 424
          F YDKSTDP++ T GHI Q+LIG +KK G YLIS G K+ITV KE N++KA S+
60  Sbjct: 393 FLYDKSTDPEQHTSGHINQDLTGKVTVKKVGCQHYLISQSGSKNITVVKEDNQLKAP--SV 450

Query: 425  FPENLVGSWKGEVEDGTVTMTFFDKDGKVTQK-KVYKDSKSKESNHSKAVTKLEDKGNGLY 483
          FPE+++G+W G+ ++ M+ DG +T K + K ++SKE+ +AK++K+EDKGNNG Y
65  Sbjct: 451 FPESILGTWTGQANGLSIHMSLASDGTITTKVEDQKGNRSKET-RTAKISKVEDKGNNGFY 509

Query: 484  LYQYESGTDTTTFV-TGGIGGLKVKYAYGIKIEGKNIIPVIWQTS SDGEFDYHKPLLSKP 542
          LY + G+D + V GG+GG VKYAYG KI G PV+WQ + EFDY KPL
70  Sbjct: 510 LYTPDPGSDISALVPEGGLGGANVKYAYGFKISGKTASPVVWQAALTHEFDYTKPLSGVT 569

```

-2277-

Query: 543 LTKQ 546
 L KQ
 Sbjct: 570 LQKQ 573

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 9065> which encodes amino acid sequence <SEQ ID 9066>. An alignment of the GAS and GBS sequences follows:

Score = 47.3 bits (110), Expect = 4e-07
 Identities = 65/303 (21%), Positives = 119/303 (38%), Gaps = 18/303 (5%)

Query: 153 FYILGIGTSISIVVALTRFVKEISLNFKKLANKMGIEVLSENEYSQII---EFDDI 209
 +YIL + T I+ +V + +S F +KKL KM + +QI EF D+
 Sbjct: 37 YYILSV-TIIACIVGGIVNLFLLSSVFTSLKCLKQKMDISQRCPTTKAQICSPQEFKDL 95

Query: 210 LRTLHIKGDNLKSLIEREILEKQDLSFQIAALSHDIKTPXXXXXXXXXXXXXXXXXXXXQE 269
 + L+S + +++ + IA LSHDIKTP +
 Sbjct: 96 ETAFNQMSSELESTFKSLNESEREKTMIAQLSHDIKTPITSIQSTVEGILDGIISEEV 155

Query: 270 GYIVSMNNSISVFEGYFNLSIYTRML-----SEDRSVKLILVEELLSELHFEVDDL 321
 Y + N+IS N L+ + +E + I +++LL ++ E +
 Sbjct: 156 NYYL---NTISRQTNRLNHLVEELSFITLETMSDTAEPHKEETIYDKLLIDILSEFQLV 212

Query: 322 LNINNIEFSICNRLIITSFYGDERNLIRALSNNLVNNAIRFMPVLDKKIEVILSESQEIQH 381
 N + I ++ + L R L NL+ NA ++ + + + I
 Sbjct: 213 FEKENRQVMIDVAPDVSKLSSQYDKLSRILLNLISNAXKYSDF-GSPLTIKAYSNRQDIV 271

Query: 382 FEIWNNGERFSDSTLKKGDKLFYTEDYSRGNK--HYGIGLAFVKGVAIKHGGNLQLNPA 439
 +I + G D L Y + SR K +G+GL + +A + G++ + +
 Sbjct: 272 IDIIDQGYGIKDEDLASIFNRLYRVSSRNMTGGHGLGLYIARQLAHQLNGDILVESQY 331

Query: 440 RGG 442
 + G
 Sbjct: 332 QKG 334

A related sequence was also identified in GAS <SEQ ID 9135> which encodes the amino acid sequence <SEQ ID 9136>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -3.56 Transmembrane 145 - 161 (145 - 164)

----- Final Results -----

bacterial membrane --- Certainty=0.2423(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

SEQ ID 6254 (GBS280) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 8; MW 63.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 7; MW 88.7kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2022

A DNA sequence (GBSx2133) was identified in *S.agalactiae* <SEQ ID 6257> which encodes the amino acid sequence <SEQ ID 6258>. This protein is predicted to be ribosomal large subunit pseudouridine synthase D (rluC). Analysis of this protein sequence reveals the following:

Possible site: 22

-2278-

>>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -4.62 Transmembrane 2 - 18 (1 - 19)

----- Final Results -----

5 bacterial membrane --- Certainty=0.2848(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB12749 GB:Z99108 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 97/251 (38%), Positives = 147/251 (57%), Gaps = 15/251 (5%)

Query: 86 KHVLIINNEFINWQTVVQENDTITLIFDDEDYPTKKIPLGRAELIDCLYEDEHLIIVNKPE 145
 + + +N+E + +V++ D + + + + + G +D L+ED H++I+NKP

15 Sbjct: 43 QQIKVNHESVLNNMIVKKGDRVFIDLQSEASSVIPEYGE---LDILFEDNHMLIINKPA 99

Query: 146 GMKTHGNQPNIEIALLNHVSAY----SGQTCYV--VHRLDMETSGAVLFAKNPFILPLINQ 199
 G+ TH N+ + L ++ AY +G+TC V VHRLD +TSGA++FAK+ +++Q

20 Sbjct: 100 GIATHPNEDGQTGTLANLIAVHYQINGETCKVRHVHRLDQDTSGAIVFAKHRLAHAILDQ 159

Query: 200 RLERKEIWREYWALVEGKFSKPHQVLRDKIGRNR-HDRRKRIIDSKNGQHAMIIDVL-- 256
 +LE+K + R Y A+ EGK K + IGR+R H R+R+ S GQ A+T V+

25 Sbjct: 160 QLEKKTLKRTYTAIAEGKLRTKKGTINPPIGRDRSHPTRRRV--SEGGQTAVTHFKVMAS 217

Query: 257 KYIQNSSLIKCRLETGRTHQIRVHLSHHGHPLIGDPLYNPSSN-NERLMLHAHRLTLSHP 315
 + SL++ LETGRTHQIRVHL+ GHPL GD LY S R LHA+++ HP

30 Sbjct: 218 NAKERLSLVELELETGRTHQIRVHLASLGHPLTGDLSLYGGGSKLLNRQALHANKVQAVHP 277

Query: 316 LTCETISVEAP 326
 +T E I EAP

Sbjct: 278 ITDELIVAEAP 288

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6259> which encodes the amino acid sequence <SEQ ID 6260>. Analysis of this protein sequence reveals the following:

35 Possible site: 38
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.4198(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 172/278 (61%), Positives = 212/278 (75%), Gaps = 2/278 (0%)

45 Query: 63 TVKELLEDFLIPRKIRHFLRVKKHVLINNEFINWQTVVQENDTITLIFDDEDYPTKKIP 122
 TVK LLE+ LIPRKIRHFLR KKHVLIN +NWQ+ V+ D + L FD EDYP K I
 Sbjct: 2 TVKALLEEQLLIPRKIRHFLRTKKHVLINGHSVNWQSCVKYGDQVKLFFDHEDYPEKIIV 61

50 Query: 123 LGRAELIDCLYEDEHLIIVNKPEGMKTHGNQPNIEIALLNHVSAYSQTCYVVHRLDMETS 182
 +G+AE + CLYEDEH+IIVNKPEGMKTHGN P E+ALLNHVSAY+GQTCYVVHRLD ETS
 Sbjct: 62 MQQAEKVTCLYEDEHIIIVNKPEGMKTHGNDPTELALLNHVSAYTGQTCYVVHRLDKETS 121

55 Query: 183 GAVLFAKNPFILPLINQRLERKEIWREYWALVEGKFSKPHQVLRDKIGRNRHRRRKRIID 242
 GA+LFAK PFILP++N+ LE+++I REY ALV G IGR+RHDRKR++D
 Sbjct: 122 GAILFAKTPFILPILNRLLKRDHREYLALVHGSLSRPRVTYHHPIGRHRHRRKRVD 181

60 Query: 243 SKNGQHAMIIDVLK-YIQNSSLIKCRLETGRTHQIRVHLSHHGHPLIGDPLY-NPSSNN 300
 NG+ A+T + ++K + + +SL+ C+L+TGRTHEQIRVHL+H GH L GDPLY N +
 Sbjct: 182 PINGKKAITEVTLVKNFHKTSALLTQQLQTGRTHQIRVHLAHQGHVLFGDPLYSGNKKDC 241

Query: 301 ERLMLHAHRLTLSHPLTCETISVEAPSSTFEKILNNYK 338
 RLMLHA++L L HPLT E I V+A S+TF+ +LN K
 Sbjct: 242 ARLMLHAYQLRLKHPLTQEDICVQAKSATFDAVLNAQK 279

-2279-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2023

- 5 A DNA sequence (GBSx2134) was identified in *S.agalactiae* <SEQ ID 6261> which encodes the amino acid sequence <SEQ ID 6262>. Analysis of this protein sequence reveals the following:

```

Possible site: 52
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -9.02      Transmembrane      98 - 114 ( 93 - 119)
10
      ----- Final Results -----
                  bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15                  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF04735 GB:AF101780 penicillin-binding protein 2a
      [Streptococcus pneumoniae]
      Identities = 424/773 (54%), Positives = 555/773 (70%), Gaps = 47/773 (6%)
20
      Query: 2   KLFQDKFIDLFRVDEDNDEMTRKNEQETREETSNLDGEEVYDIDDITRPSKSQYQRGIRHQ 61
                  KLF+KF+ LF+ +ETS L+ + I R S+S
      Sbjct: 5   KLFQKFLSLFK-----KETSLEDS-----STILRRSRS----- 34

25      Query: 62  KENAKSRPEWLQKVDRLPSPKNPIRRFWRRYRIGKLLFIALMAFILIFGSYLFYLSKTA 121
                  DR + PIR+FWRRY + K++ I ++ L+ G YLF ++K+
      Sbjct: 35  -----DRKKLAQVGPIRKFWRRYHLTKIILGLSAGLLVGIYLFVAVAKST 80

30      Query: 122 TVSDLQSAKLTITTIYDKNKEYAGKLSGQKGTVELNAISDHLKNAVIATEDRTFYENNG 181
                  V+DLQ+ALKT T I+D+ ++ AG LSGQKGTVEL IS +L+NAVIATEDR+FY+N+G
      Sbjct: 81  NVNDLQNALKTRTLIFDREEKEAGALSGQKGTVELTDISKNLQNAVIATEDRSFYKNDG 140

      Query: 182  VNFKRFFLAIVTAGRSGGGSTITQQLAKNAYLSQDQTIKRKAREFFLALELTKKYSKAEI 241
                  +N+ RFFLA+ T G+ GGGSTITQQLAKNAYLSQDQ++RKA+EFFLALEL+KKYSK +I
35      Sbjct: 141 INYGRFFLAIVTAGRSGGGSTITQQLAKNAYLSQDQTVRKAKEFFLALELSKKYSKEQI 200

      Query: 242  LTMYLNNYSYFGNGVWGVEDASRKYFGTSAANLTVDEAATLAGMLKGPEVYNPYYSVENAT 301
                  LTMYLNN+YFGNGVWGVEDAS+KYFG SA+ +++D+AATLAGMLKGPE+YNP SVE++T
40      Sbjct: 201 LTMYLNNAYFGNGVWGVEDASKYFGVSASEVSLDQAATLAGMLKGPELYNPLNSVEDST 260

      Query: 302  NRRDVTIAAMVDAGKLTKSQAKEAASIGMKNRLADTYAGKINDYRYPYFDAVVNEAIDT 361
                  NRRDVTLV MV AG + K+Q EAA + M ++L D Y GKI+DYRYPYFDAVVNEA+
      Sbjct: 261 NRRDVTLQNMVAAGYIDKNQETEAAEVDMTSQLHDKYEGKISDYRYPYFDAVVNEAVSK 320

45      Query: 362  YGISEKDIVNNGYKIYALDQNYQSGMQKTFFDTSFLFPVSDYDQSAQGASVALDPKTGG 421
                  Y ++E++IVNNGY+IYT LDQNYQ+ MQ +++TSLFP ++ DG AQ SVAL+PKTGG
      Sbjct: 321 YNLTEBIVNNGYRIYTELDQNYQANMQIVYENTSLFPRAE-DGTFAQSGSVALEPKTGG 379

      Query: 422  VRGLVGRVQSTKDAQFRSFNYATQSKRSPASTIKPLVVYSPAIASGWSIDKELENKVQDF 481
                  VRG+VG+V FR+FNATQSKRSP STIKPLVVY+PA+ +GW+++K+L N +
50      Sbjct: 380 VRGVVQVADNDKTGFRNFNYATQSKRSPGSTIKPLVVYTPAVEAGWALNKQLDNHTMQY 439

      Query: 482  HGYPKPSNYGGIET-ESIPMYQALANSYNIPAVYTLDKLGINKAFTYGRKFGFLNMSSANKE 540
                  YK NY GI+T +PMYQ+LA S N+PAV T++ LG++KAF G KFGLNM ++
55      Sbjct: 440 DSYKVDNYAGIKTSREVPMYQSLAESLNLPAVATVNDLGVDKAFEAGEKFGLNMEKVDRV 499

      Query: 541  LGVALGGSVITNPLEMAQAYSTFANDGIMHRAHLITRIETANGKLVKQFTDKPKRVISRS 600
                  LGVALG V TNPL+MAQAY+ FAN+G+M AH I+RIE A+G+++ + KRVI +S
      Sbjct: 500 LGVALGSGVETNPLQMAQAYAAFANEGLMPEAHFISRIENASQGVIAASHKNSQKRVIDKS 559

60      Query: 601 VASKMTSMMLGTFFSNGTAINANVGYTMAAGKTGTTFETDFNPNLSGDQWVVGYPDPVVISQ 660
                  VA KMTSMMLGTFF+NGT I+++ Y MAGKTGTTE FNP + DQWV+GYTPDVVIS

```

-2280-

Sbjct: 560 VADKMTSMMLGTFNNGTGISSSPADYVMAGKTGTTEAVFNPEYTSQWVIGYTPDVVISH 619

Query: 661 WVGFKNTDKHHYLDSSAGTASNIFSTQASYILPYTKGSSFTTHIENAYFQNGIGSVYNAQ 720

W+GF TD++HYL S++ A+++F A+ ILPYT GS+FT +ENAY QNGI +

Sbjct: 620 WLGFPPTDENHYLAGSTSNAAHVFRNIANTILPYTPGSTFT-VENAYKQNGIAPANTKR 678

Query: 721 DASNTTNQESRSIINDLKDSASKAAQDISRAVEDSNFQEKVKDAWNSLKDYFR 773

N ++ ++D++ A + SRA+ D+ +EK + W+S+ + FR

Sbjct: 679 QVQTNDNSQTDNLSDIRGRAQSLVDEASRAISDAKIKEKAQTIWDSIVNLFR 731

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6263> which encodes the amino acid sequence <SEQ ID 6264>. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -7.96 Transmembrane 104 - 120 (99 - 124)

----- Final Results -----

bacterial membrane --- Certainty=0.4185(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF04735 GB:AF101780 penicillin-binding protein 2a [Streptococcus pneumoniae]

Identities = 414/730 (56%), Positives = 539/730 (73%), Gaps = 17/730 (2%)

Query: 50 TKNSEQDPATALQRSRAYEGSPKSRPAWLQKLEAVLPSQPPIRRFWRRYHIGKLLMILI 109

T E +T L+RSR+ +KL V PIR+FWRRYH+ K+++IL

Sbjct: 18 TSELESDSTILRRSRSDR-----KKLAQV-----GPIRKFWRRYHLTKIILILG 62

Query: 110 GTLVLLLSYLYFYSKTAKVSDLODALKATTVIYDHKGEYAGSLSGQKGSYVELNAISDD 169

+ LL+G YLF ++K+ V+DLQ+ALK T+I+D + + AG+LSGQKG+YVEL IS +

Sbjct: 63 LSAGLLVGIYLFVAVAKSTNVNDLQNALKTRTLIFDREEKEAGALSGQKGTVELTDISK 122

Query: 170 LENAIVATEDRTFYNSNGINLKRFLAVVTAGRFGGGSTITQQLAKNAYLSQDQTIKRKA 229

L+NAIVATEDR+FY N GIN RF LA+VTAGR GGGSTITQQLAKNAYLSQDQ++RKA

Sbjct: 123 LQNAIVATEDRSFYKNDGINYGRFFLAIVTAGRSGGGSTITQQLAKNAYLSQDQTVRKA 182

Query: 230 REFFLALELTKKYSKKDILTMYLNNYFCNGVWGVEDASQKYFGTTAANLTLDEAATLAG 289

+EFFLALEL+KKYSK+ ILTMYLNN+YFCNGVWGVEDAS+KYFG +A+ ++LD+AATLAG

Sbjct: 183 KEFFLALELSKKYSKEQILTMYLNNAYFCNGVWGVEDASKKYFGVSASEVSLDQAATLAG 242

Query: 290 MLKGPEIYNPYHSLKNATHRRDVTVLGAMVDAKKITQTKAQARAVGLKNRLADTYVGKTD 349

MLKGPE+YNP +S++++T+RRDVTL MV A I + + +A V + ++L D Y GK

Sbjct: 243 MLKGPELYNPLNSVEDSTNRRDVTVLQNMVAAGYIDKNQETEAEEVDMTSQLHDKYEGKIS 302

Query: 350 DYKYPYSYFDAVISEAIATYGLSEKDIVNNGYKVYTELDQNYQTGMQTTFNDELFPVSAY 409

DY+YPSYFDAV++EA++ Y L+E++IVNNGY++YTELDQNYQ MQ + N LFP A

Sbjct: 303 DYRYPSYFDAVNEAVSKYNLTBBEIVNNGYRIYTELDQNYQANMQIVYENTSLFP-RAE 361

Query: 410 DGSSAQAAASVALDPKTGGVRGLGRVNSSENPTFRSFNYATQAKRSPASTIKPLVVYAPA 469

DG+ AQ+ SVAL+PKTGGVRG++G+V ++ FR+FNATQ+KRSP STIKPLVVY PA

Sbjct: 362 DGTFAQSGSVALEPKTGGVRGVVQVADNDKTGFRNFNYATQSKRSPGSTIKPLVVYTPA 421

Query: 470 VASGWSIEKELENTVQDFDGYQPHNY-GNYESEDVPMYQALANSYNIPAVSTLNDIGIDK 528

V +GW++ K+L N +D Y+ NY G S +VPMYQ+LA S N+PAV+T+ND+G+DK

Sbjct: 422 VRAGWALNKQLDNHTMQYDSYKVDNYAGIKTSREVPMYQSLAESLNLPAVATVNDLGVDK 481

Query: 529 AFTYTGKTFGLDMSSAKKELGVALGGSVTTNPLEMAQAYAAAFANNGVIHPAHLINRIENAR 588

AF G+ FGL+M + LGVALG V TNPL+MAQAYAAAFAN G++ AH I+RIENA

Sbjct: 482 AFEAGEKFGLNMEKVDRLVGLVALGSGVETNPLQMAQAYAAAFANGLMPEAHFISRIENAS 541

Query: 589 GEVLKTFDTAKRVVSQSVDKMTAMMLGTFNNGTAVNANVGYTLAGKTGTETNFPND 648

G+V+ + + KRV+ +SVADKMT+MMLGTF+NGT +++++ Y +AGKTGTTE FNP+

Sbjct: 542 GQVIASHKNSQKRVIDKSVADKMTSMMLGTFNNGTGISSSPADYVMAGKTGTTEAVFNPE 601

-2281-

Query: 649 LAGDQWVIGYTPDVVISQWVGFNQTDENHYLTDSSAGTASAIFFSTQASYILPYTKGSQFH 708
 DQWVIGYTPDVVIS W+GF TDENHYL S++ A+ +F A+ ILPYT GS F
 Sbjet: 602 YTSQWVIGYTPDVVISHLGFPTTDENHYLAGSTSNGAHVFRNIANTILPYTPGSTFT 661

Query: 709 VDNAYAQNNGISAVYGVNETGNQSGVDTSIIDGLRKSAQEASQSLSKAVDQSGLRDKAQS 768
 V+NAY QNGI+ + T + +R AQ S+A+ + +++KAQ+
 Sbjet: 662 VENAYKQNGIAPANTKRQVQTNDNSQTDDNLSDIRGAQSLVDEASRAISDAKIKEKAQT 721

Query: 769 IWKEIVDYFR 778
 IW IV+ FR
 Sbjet: 722 IWDSIVNLF 731

An alignment of the GAS and GBS proteins is shown below.

Identities = 530/715 (74%), Positives = 623/715 (87%), Gaps = 1/715 (0%)

Query: 59 RHQKENAKSRPEWLQKVDRYLPSKPNPIRRFWRRYRIGKLLFIALMAFILIFGSLFYLS 118
 R + + KSRP WLQK++ LSP+ PIRRFWRRY IGKLL I + +L+ GSYLFYLS
 Sbjet: 65 RAYEGSPKSRPAWLQKLEAVLPSQRPPIRRFWRRYHIGKLLMILIGTLVLLGSLFYLS 124

Query: 119 KTATVSDQLSALKITTTIYDKNKEYAGKLSGQKGTVELNAISDHLKNAVIATEDRTFY 178
 KTA VSDLQ ALK TT IYD EYAG LSGQKG+YVELNAISD L+NAVIATEDRTFY
 Sbjet: 125 KTAKVSQDALKATTVIYDHKGEYAGSLSGQKGSYVELNAISDDLENAVIATEDRTFY 184

Query: 179 NNGVNFKRFFLAVATLGKFGGGSTITQQLAKNAYLSQDQTIKRKAREFFLALELTKKYSK 238
 N+G+N KRF LAV T G+FGGGSTITQQLAKNAYLSQDQTIKRKAREFFLALELTKKYSK
 Sbjet: 185 NSGINLKRFLAVVTAGRFGGGSTITQQLAKNAYLSQDQTIKRKAREFFLALELTKKYSK 244

Query: 239 AEILTMYLNNSYFGNGVGVGEDASRKYFGTSAANLTVDEAATLAGMLKGPEVYNPYYSVE 298
 +ILTMYLNNSYFGNGVGVGEDAS+KYFGT+AAANL+DEAATLAGMLKGPE+YNPY+S++
 Sbjet: 245 KDILTMYLNNSYFGNGVGVGEDASQKYFGTTAANLTLDEAATLAGMLKGPEIYNPYHSLK 304

Query: 299 NATNRRDRTLAAVMDAGKLTQSQAEEASIGMKNRLADTYAGKINDYRPSYFDDAVVNEA 358
 NAT+RRDRTL AMVDA K+T+++A++A ++G+KNRLADTY GK +DY+YPSYFDDAV++EA
 Sbjet: 305 NATHRRDRTLGAAMDAAKITQTQAQARAVGLKNRLADTYVGKTDYKPSYFDDAVISEA 364

Query: 359 IDTYGISEKDIVNNGYKIYALDQNYQSGMQKTFDDTSLFPVSDYDQSAQAGSVALDPK 418
 I TYG+SEKDIVNNGYK+YT LDQNYQ+GMQ TF++ LFPVS YDG SAQ ASVALDPK
 Sbjet: 365 IATYGLSEKDIVNNGYKVYTELDQNYQTMQTTFNDELFPVSAYDGSQAASVALDPK 424

Query: 419 TGGVRGLVGRVQSTKDAQFRSFNYATQSKRSPASTIKPLVVYSPAIASGWSIDKELPNKV 478
 TGGVRGL+GRV S+++ FRSFNYATQ+KRSPASTIKPLVVY+PA+ASGWSI+KELPN V
 Sbjet: 425 TGGVRGLIGRVNSENPTFRSFNYATQAKRSPASTIKPLVVYAPAVASGWSIEKELPNTV 484

Query: 479 QDFHGYKPSNYGGIETESIPMYQALANSYNIPAVYTLDKLGINKAFTYGRKFGLNMSSAN 538
 QDF GY+P NYG E+E +PMYQALANSYNIPAV TL+ +GI+KAFTY+ FGL+MSSA
 Sbjet: 485 QDFDGYQPHNYGNYESEDVPMYQALANSYNIPAVSTLNDIGIDKAFTYGKTFGLDMSSAK 544

Query: 539 KELGVALGGSVTTNPLEMAQAYSTFANDGIMHRAHLITRIETANGKLVKQFTDKPKRVIS 598
 KELGVALGGSVTTNPLEMAQAY+ FAN+G++H AHLI RIE A G+++K FIDK KRV+S
 Sbjet: 545 KELGVALGGSVTTNPLEMAQAYAAFANNGVIHPAHLINRIENARGEVLKTFIDKAKRVVS 604

Query: 599 RSVASKMTSMMLGTFSNGTAINANVYGYTMAGKTGTTETDFNPNLSGDQWVGYTPDVVI 658
 +SVA KMT+MMLGTFSNGTA+NANVYGYT+AGKTGTTET+FN+L+GDQWV+GYTPDVVI
 Sbjet: 605 QSVADKMTAMMLGTFSNGTAVNANVYGYTLAGKTGTTETNFPDLAGDQWVIGYTPDVVI 664

Query: 659 SQWVGFKNTDKHHYLTDSAGTASNIFSTQASYILPYTKGSFTHIENAYFQNGIGSVYN 718
 SQWVG F TD++HYLTDSAGTAS IFSTQASYILPYTKGS F H++NAY QNGI +VY
 Sbjet: 665 SQWVGFNQTDENHYLTDSSAGTASAIFFSTQASYILPYTKGSQF-HVDNAYAQNNGISAVYG 723

Query: 719 AQDASNTTNQESRSIINDLKDSASKAAQDISRAVEDSNFQEKVKDAWNSLKDYFR 773
 + N + +++SII+ L+ SA +A+Q +S+AV+ S ++K + W + DYFR
 Sbjet: 724 VNETGNQSGVDTSIIDGLRKSAQEASQSLSKAVDQSGLRDKAQSIIWKEIVDYFR 778

SEQ ID 6262 (GBS397d) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 153 (lane 13; MW 76kDa) and in Figure 184 (lane 9; MW 76kDa).

A DNA sequence (GBSx2135) was identified in *S.agalactiae* <SEQ ID 6265> which encodes the amino acid sequence <SEQ ID 6266>. This protein is predicted to be M-like protein. Analysis of this protein sequence reveals the following:

```

Possible site: 27
>>> Seems to have no N-terminal signal sequence
    INTEGRAL      Likelihood = -10.56      Transmembrane  609 - 625 ( 599 - 628)
10    INTEGRAL      Likelihood = -0.00      Transmembrane  19 - 35 ( 19 - 35)

----- Final Results -----
          bacterial membrane --- Certainty=0.5225(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

>GP:CAB91647/ GB:AJ130830 cell wall protein, putative [Zea mays]
Identities = 106/182 (58%), Positives = 123/182 (67%), Gaps = 8/182 (4%)

20 Query: 396 KEDKKPDVKPEAKPEAK--PDVKPEAKPDVKPEAKPDVKPEAKPDV--KPEA 451
K + KP+ KPE KPE K P KPE KP+ KPE KP+ KPE KP KPE KP+ KPE
Sbjct: 116 KPEPKPEPKPEPKPEPKIKPKPKPEPKPEPKPEHKPEPKPEPKPKPEPKPEPQPKPEP 175

25 Query: 452 KPDVKPKAKPDVKPEA--KPDVKPDVKPDVKPEA--KPEDKPDVKEDVKPEAKPDVKPEA 507
KP+ KP+ KP+ KPE KP+ KP+ KP+ KPE KPE KP+ KP+ KPE KP+ KPE
Sbjct: 176 KPEPKPEPKPEPKPEPQPKPEPKPEPKPEPKPEPQPKPEPKPEPKPEPKPEPKPEPKPEP 235

30 Query: 508 KPEAKPEAKPEAKPEAKPEAKPDVKPEAKPDVKPEAKPEAKPEAKSEAKPEAKLEAKPEA 567
KPE KPE +PE KPE KPE KP P+ +P KPE KPE KPE K E KPE K E KPE
Sbjct: 236 KPEPKPEPKPEPKPEPKPEPKPKPKPEPQPKPEPKPEPKPEPKPEPKPEPKPEPKPEPKPEP 295

35 Query: 568 KP 569
KP
Sbjct: 296 KP 297

A related GBS gene <SEQ ID 8957> and protein <SEQ ID 8958> were also identified. Analysis of this protein sequence reveals the following:

```

40      Lipop: Possible site: -1      Crend: 8
      McG: Discrim Score:      -5.20
      GvH: Signal Score (-7.5): 3.07
          Possible site: 27
      >>> Seems to have no N-terminal signal sequence
45      ALOM program      count: 2 value: -10.56 threshold: 0.0
          INTEGRAL      Likelihood ==-10.56      Transmembrane      609 - 625 ( 599 - 628)
          INTEGRAL      Likelihood = -0.00      Transmembrane      19 - 35 ( 19 - 35)
          PERIPHERAL      Likelihood = 8.54      139
      modified ALOM score: 2.61
50
      *** Reasoning Step: 3

      ----- Final Results -----
55          bacterial membrane --- Certainty=0.5225(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

      LPXTG motif: 596-600

```

5

10

15

20

25

35

40

45

50

55

-2284-

Example 2025

A DNA sequence (GBSx2136) was identified in *S.agalactiae* <SEQ ID 6267> which encodes the amino acid sequence <SEQ ID 6268>. This protein is predicted to be transcription antitermination protein nusg (nusG). Analysis of this protein sequence reveals the following:

```

5      Possible site: 48
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3203 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15      >GP:CAA53738 GB:X76134 nusG [Staphylococcus carnosus]
      Identities = 90/175 (51%), Positives = 118/175 (67%), Gaps = 2/175 (1%)

      Query: 7   KGWFLVLTQYSGYENKVKENLLQRAQTYNMLDNILRVEIPTQTVNVEKNGKTKEIENRFP 66
                K W+ + TYSGYENKVK+NL +R ++ NM + I RV IP +      K+GK K++ + FP
20      Sbjct: 8   KRWYAVHTYSGYENKVKKNLEKRVESMMNTEQIFRVVPIPEEETQVKDGKAKKLTKKTFP 67

      Query: 67  GYVLVEMVMTDEAWFVVRNTPNVTGFGVSHGNRSKPTPLLEEIRISILISMGTVDVFD 126
                GYVLVE+VMTDE+W+VVRNTP VTGFGVS G SKP PLL +E+R IL MG D
25      Sbjct: 68  GYVLVELVMTDESWYVVRNTPGVTFGVSAGAGSKPNPLLPDEVRFILKQGMKEKTIDV 127

      Query: 127 NIKEGDVVQIIDGAFIQEGRVVEIENNKVKL--MINMFGSETQAELELYQVAEL 179
                ++ G+ V+I G F Q G V EIE +K KL +++MFG ET E+E Q+ +L
30      Sbjct: 128 EVEVGQVRIKSGPFANQVGEVQIEADKFKLTVLDMFGRETPEVEVEFDQIEKL 182

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6269> which encodes the amino acid sequence <SEQ ID 6270>. Analysis of this protein sequence reveals the following:

```

      Possible site: 53
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
35      bacterial cytoplasm --- Certainty=0.3874 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

40      Identities = 170/179 (94%), Positives = 178/179 (98%)

      Query: 1   MLDSFDKGFVLQYSGYENKVKENLLQRAQTYNMLDNILRVEIPTQTVNVEKNGKTKEI 60
                MLDSFDKGFVLQYSGYENKVKENLLQRAQTYNMLDNILRVEIPTQTVNVEKNG+TKEI
45      Sbjct: 6   MLDSFDKGFVLQYSGYENKVKENLLQRAQTYNMLDNILRVEIPTQTVNVEKNGQTKEI 65

      Query: 61  EENRFPGYVLVEMVMTDEAWFVVRNTPNVTGFGVSHGNRSKPTPLLEEIRISILISMGT 120
                EENRFPGYVLVEMVMTDEAWFVVRNTPNVTGFGVSHGNRSKPTPLLEEIR+IL+SMGT
50      Sbjct: 66  EENRFPGYVLVEMVMTDEAWFVVRNTPNVTGFGVSHGNRSKPTPLLEEIRAILLSMGQT 125

      Query: 121 VDVFDTNIKEGDVVQIIDGAFIQEGRVVEIENNKVKLMINMFGSETQAELELYQVAEL 179
                +DVFDTNIKEGDVVQIIDGAF+GQEGRVVEIENNKVKLM+NMFGSET AE+ELYQ+AEL
55      Sbjct: 126 IDVFDTNIKEGDVVQIIDGAFMGQEGRVVEIENNKVKLMINMFGSETVAEVELYQIAEL 184

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2285-

Example 2026

A DNA sequence (GBSx2137) was identified in *S.agalactiae* <SEQ ID 6271> which encodes the amino acid sequence <SEQ ID 6272>. This protein is predicted to be a glycosyl transferase. Analysis of this protein sequence reveals the following:

```

5      Possible site: 16
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.1558(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15      >GP:AAF28363 GB:AF224467 putative glycosyl transferase [Haemophilus
      ducreyi]
      Identities = 98/259 (37%), Positives = 155/259 (59%), Gaps = 10/259 (3%)

      Query: 5 VALAVDSNYLDKALVTIKSICVYNRNITFYLFNQDTPVEWVRNINRKLEPLGSKLINVKI 64
      + LA + +Y + L TIKSI ++N++I FYL N+D P EW +N KL L S++I++K+
20      Sbjct: 10 IVLAANQSYSEYILTTIKSIYLNHKKHIRFYLLNRDYPTEWFDILNNKLRKLNSEIIDIKV 69

      Query: 65 YNYDIAHLTFTFLTVS---TWFRFLADYIPSSRVLYLDSDIIVNTNLDYLFELDFKGYL 121
      N I + T+ +S T+FR F++D+I +V+YLD+DI+VN +L L++ D Y+L
25      Sbjct: 70 TNDTIKNFKTYSHISSDITFFRYFISDFIEQDKVIYLDADIVVNGSLTELYQTDISNYFL 129

      Query: 122 AAVKDPHKNE----EGGFNAGMLLANLELWREDGLTKTLKTAELHRVVKITGDQSILNI 177
      AAVKD + FNAGMLL N + WRE +T+ L +E+ + DQSILN+
      Sbjct: 130 AAVKDIISEKIYVMNHIFNAGMLLINNKKWREHNITQFCLSLSEKYINSLPDADQSILNL 189

30      Query: 178 VCHNRWLSLNKTWNF--QTYDVVSRYNHRSYLYLNINENRTPNIIHFLTSDKPNWENSVAR 235
      + ++WL LN+ +N+ T + +Y YL ++ P IIH+ T KPW R
      Sbjct: 190 IFKDKWLKLNRGYNYLIGTDYLFYFKYCKTRYLE-DLGETIPLIIHYNTEAKPWLNIFNTR 248

35      Query: 236 FRELWYYFQQLDFCQLTGK 254
      FR ++W+Y++L++ + K
      Sbjct: 249 FRNIYWFYVELNWQDIYAK 267

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2027

A DNA sequence (GBSx2138) was identified in *S.agalactiae* <SEQ ID 6273> which encodes the amino acid sequence <SEQ ID 6274>. Analysis of this protein sequence reveals the following:

```

45      Possible site: 20
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
50      bacterial cytoplasm --- Certainty=0.0417(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2286-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2028

- 5 A DNA sequence (GBSx2139) was identified in *S.agalactiae* <SEQ ID 6275> which encodes the amino acid sequence <SEQ ID 6276>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -2.60      Transmembrane      306 - 322 ( 306 - 322)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

- 15 The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF28363 GB:AF224467 putative glycosyl transferase [Haemophilus
      ducreyi]
      Identities = 88/259 (33%), Positives = 156/259 (59%), Gaps = 11/259 (4%)

20  Query: 7   VVLADYSYIRQIETTLKSLCVYHENLSIFIFNQDIPQEWFLAMKDRVGQTGNQIQDVKL 66
      +VLA + SY I TT+KS+ ++++++ ++ N+D P EWF + +++ + ++I D+K+
      Sbjct: 10 IVLAANQSYSEYILTTIKSIYLNKHIRFYLLNRDYPTEWFDILNNKLRKLNSEIIDIKV 69

      Query: 67 FHDHLSPKWENKLNHINY-MTYARYFIPQYISADTVLYLSDLVVTTINLNDNLFQISLDN 125
      +D + K +HI+ T+ RYFI +I D V+YLD+D+VV +L L+Q + N
      Sbjct: 70 TNDTIK---NFKTYSHISSDTFFFRYFISDFIEQDKVIYLDADIVVNGSLTELYQTDISN 126

      Query: 126 AYLAAVP-----ALFGLGYGFNAGVMVINNRWRQENMTIKLIEKNQKEIENANEGDQTI 180
      +LAAV ++ + FNAG+++INN++WR+ N+T + ++K I + + DQ+I
      Sbjct: 127 YFLAAVKDIISEKIYVNNHIFNAGMLLINNKKWREHNITQFCLSLSEKYINSLPDADQSI 186

      Query: 181 LNRMFENQVIYLD DTYNFOIGFD-MGAAIDGHKFIFDIPITPLPKIIHYISGKIPWQTL 239
      LN +F+++ + L+ YN+ IG D + +++ D+ T +P ITHY + KPW +
      Sbjct: 187 LNLIFKDKWLKLNRGYNYLIGTDYLFYKYGKTRYLEDLGET-IPLIIHYNTEAKPWLNIF 245

35  Query: 240 NMRLREVWVWHYNLLEWSSI 258
      N R R ++W Y L W I
      Sbjct: 246 NTRFRNIYWFFYELNWQDI 264
```

- 40 No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 6276 (GBS395) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 5; MW 47.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 83 (lane 8; MW 72kDa) and in Figure 177 (lane 5; MW 72kDa).

- 45 GBS395-GST was purified as shown in Figure 217, lane 7.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2029

- 50 A DNA sequence (GBSx2140) was identified in *S.agalactiae* <SEQ ID 6277> which encodes the amino acid sequence <SEQ ID 6278>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
```


-2287-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1633(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2030

A DNA sequence (GBSx2141) was identified in *S.agalactiae* <SEQ ID 6279> which encodes the amino acid sequence <SEQ ID 6280>. Analysis of this protein sequence reveals the following:

15 Possible site: 44
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.16 Transmembrane 36 - 52 (36 - 52)
 ----- Final Results -----
 20 bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10243> which encodes amino acid sequence <SEQ ID 10244> was also identified.

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC77330 GB:AE000508 orf, hypothetical protein [Escherichia coli K12]
 Identities = 75/260 (28%), Positives = 123/260 (46%), Gaps = 22/260 (8%)
 30 Query: 6 VGLVLEGGGMRLGYTAGVLDLADAGIK-IDGIVSVSAGALFGVNFVSRQRERALRYNKK 64
 + LV EGGG RG++TAGVLD F+ A D + SAGA F+ Q A + +
 Sbjct: 25 IALVCEGGGQRGIFTAGVLDFMRAQFNPFDLYLGTSAQAQLSAFICNPGYARKVIMR 84
 Query: 65 YLSHPKMYSLRSWFRTGNFVNKDF----TYEVEPMKLD----VFDDEAFKSSIDFYVVA 116
 Y + ++ + R GN ++ D+ T ++P+++D +FD S FY+ A
 35 Sbjct: 85 YTTKREFFDPLRFVRGGNLIDLWLVEATASQMPQLQMDTAARLFD-----SGKSFYMCA 138
 Query: 117 TEMTSGKPEYFKIDSVEFQMEILRASSALPVVSKM-VDWQGGKKYLDGGLSDSIPVDFARG 175
 P YF + + ++++RASSA+P + V +G YLDGG+SD+IPV A
 40 Sbjct: 139 CRQDDYAPNYF-LPTKQNWLDVIRASSAIPGFYRSGVSLEGINYLDGGISDAIPVKEAR 197
 Query: 176 LGFDKLIIVMTRPLNYQKKPSSGR-----LYKTLYRKYPNFVKTASNRYQQYNNLSLEKVM 230
 G L+V+ T P P + L + + N V+ Y+ +EK
 Sbjct: 198 QGAKTLLVIRTVP SQMYTPOWFKRMRWLGDSSLQPLVNLVQHETSRYDIIQQFIEKPP 257
 45 Query: 231 SLEKTGDLFAIRPSKSLVIG 250
 + +++ +P S+ +G
 Sbjct: 258 GKLRIFEIYPPKPLHSLALG 277

No corresponding DNA sequence was identified in *S.pyogenes*.

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8959> and protein <SEQ ID 8960> were also identified. Analysis of this protein sequence reveals the following:

-2288-

```

Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      -5.16
GvH: Signal Score (-7.5): -2.17
    Possible site: 44
5  >>> Seems to have no N-terminal signal sequence
    ALOM program   count: 1 value: -0.16 threshold: 0.0
        INTEGRAL   Likelihood = -0.16   Transmembrane 36 - 52 ( 36 - 52)
        PERIPHERAL Likelihood = 4.14     18
10    modified ALOM score: 0.53

*** Reasoning Step: 3

----- Final Results -----
15    bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

20    ORF01611(316 - 1050 of 1449)
        OMNI|NT01EC5264(37 - 289 of 369) hypothetical protein
        %Match = 9.2
        %Identity = 29.7 %Similarity = 49.8
        Matches = 74 Mismatches = 118 Conservative Sub.s = 50

25    273      303      333      363      393      420      450      480
        QKKQLYFAIL*SNINIRK*LPMLSVGLVLEGGGMRLGYTAGVLDALFDAGIK-IDGIVSVSAGALFGVNFVSRQRERALR
          : || |||| ||::||||| |: | : ||| : | :
        VGQRIPVTLGNIAPLSLRPFQPGRIALVCEGGGQRGIFTAGVLDEFMRAQFNPFDLYLGTSAQAQNLSAFICNQPGYARK
          30      40      50      60      70      80      90

30    510      540      588      618      648      678      708
        YNKKYLSHPKYMSLRSWFRITGNFVNKDF----TYEVPMKLDVFDDEAFKSSIDFYVVVATEMTSGKPEYFKIDSVEEQM
          : | : : : : | |::: | : :::: | : | ||: | | | : : :
        VIMRYTTKREFDPLRFVRGGLNLDLWLVEATASQMLQMDT--AARLFDGSKSFYMCACRQDDYAPNYF-LPTKQNWNL
35          110      120      130      140      150      160

        738      765      795      825      855      885      912      930
        EILRASSALPVPVSKM-VDWQGGKYLDGGLSDSIPVDFARGLGFDKLIIVMTRPLNYQKKPSS-GRLYKTL----YRKYPN
          :::|||||: | : | :| ||||:|:| | | | :|: | | : : | : : |
        DVIRASSAIPGFIYRSGVSLGGINYLDGGISDAIPVKEAARQGAKTLLVIRTVPSPQMYTTPQWFKRMRWLGDSSSLQPLVN
40          180      190      200      210      220      230      240

        960      990      1020      1050      1080      1110      1140      1170
        FVKTASNRYQQYNNLSLEKVMSEKLTGDLFAIREPSKSLVIGRLEKNPDKLDLSIYQLGMKDAKSVMPELNSYLMK*RKQYFS
        :|: | : : || : : : :| | : :|
        LVQHHTSYRDIQQFIEKPPGKLRIFEIYPPKPLHSIALGSRIPALREDYKLGRLCGRYFLATVGKLLTEKAPLTRHLVP
45          260      270      280      290      300      310      320

```

SEQ ID 8960 (GBS394) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 4; MW 34.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 83 (lane 7; MW 60kDa).

GBS394-GST was purified as shown in Figure 217, lane 6.

Example 2031

A DNA sequence (GBSx2142) was identified in *S.agalactiae* <SEQ ID 6281> which encodes the amino acid sequence <SEQ ID 6282>. Analysis of this protein sequence reveals the following:

```

55    Possible site: 21
        >>> Seems to have no N-terminal signal sequence

        ----- Final Results -----
60    bacterial cytoplasm --- Certainty=0.3004(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

```

-2289-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2032

A DNA sequence (GBSx2143) was identified in *S.agalactiae* <SEQ ID 6283> which encodes the amino acid sequence <SEQ ID 6284>. This protein is predicted to be transporter protein. Analysis of this protein
10 sequence reveals the following:

Possible site: 49
>>> Seems to have a cleavable N-term signal seq.

15	INTEGRAL	Likelihood = -6.85	Transmembrane	373 - 389 (370 - 395)
	INTEGRAL	Likelihood = -6.74	Transmembrane	168 - 184 (162 - 187)
	INTEGRAL	Likelihood = -6.32	Transmembrane	259 - 275 (257 - 280)
	INTEGRAL	Likelihood = -4.78	Transmembrane	286 - 302 (285 - 306)
	INTEGRAL	Likelihood = -3.19	Transmembrane	55 - 71 (54 - 71)
	INTEGRAL	Likelihood = -2.97	Transmembrane	84 - 100 (79 - 101)
	INTEGRAL	Likelihood = -2.87	Transmembrane	311 - 327 (310 - 328)
20	INTEGRAL	Likelihood = -1.44	Transmembrane	355 - 371 (355 - 371)
	INTEGRAL	Likelihood = -0.64	Transmembrane	108 - 124 (108 - 125)

----- Final Results -----

25	bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
	bacterial outside --- Certainty=0.0000(Not Clear) < succ>
	bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC22759 GB:U32790 transporter protein [Haemophilus influenzae
30 Rd]
Identities = 139/391 (35%), Positives = 221/391 (55%), Gaps = 4/391 (1%)

Query: 6 INKNNWRALIAAIVASGTDDLNIMFLAFSMSTIITDLHLSAAQAGWIGTITNLGMLVGGGL 65
+N W+ALI + V G D +++ L F +S I DL+L+ AQ G + T T +G + GG+

35 Sbjct: 5 VNSYQWKALIGSAVGYGMDGFDLLILGFMLSAISADLNLTPAQGGSLVTWTLIGAVFGGI 64

Query: 66 IFGLLADRYNKFVKFWTILIFSATGLVFFTTNLSYLYIMRFIAGIGVGGEYGI AIAIM 125
+FG L+D+Y + +V WTIL+F++ TGL L I R IAGIG+GGE+GI +A+

40 Sbjct: 65 LFGALSDKYGRVRVLTWTLILFAVFTGLCAIAQGYWDLIIYRTIAGIGLGGEGFGIGMALA 124

Query: 126 AGIVPTNKMGRISLNGIAGQVGSISSALLAGWLAPALGWRGLFLFGLLPFIVLVLMQFA 185
A P + +S + QVG + +ALL L P +GWRG+FL G+ P + +++

45 Sbjct: 125 ABAWPARHRAKAASYVALGWQVGVGAALLTPLLLPHIGWRGMFLVGIFPAFVAVFLRSH 184

Query: 186 VDDKDILDQYNTDADDEPLDI----SIKALFDTPLATQSLALMVMTTVQIAGYFGMMNW 241
++ +I Q T + S + L + SL ++V+T+VQ GY+G+M W

50 Sbjct: 185 LHEPEIFTQKQTALSTQSSFTDKLRSPQLLIKDKATSKISLGIVLTSTVQNFYGYGIMIW 244

Query: 242 LPTIIQTNLNVSVKNSSLWMIATILGMCLGMLVFGQLLDKFGPRLVYGCFLSSAICVYL 301
LP + L S+ S LW T+ GM G+ +FGQL D+ G + + F L + I + +

55 Sbjct: 245 LPNFLSKQLGFSITKSLWTAVTVCGMMAGIWFQQLADRIGRKPSFLLFQLGAVISIVV 304

Query: 302 FQFATTPMSMIIGGAVVGVFFVNGMFAGYGAMITRLYPHHIRSTANNLILNVGRAIGGFSS 361
+ T M++ GA +G FVNGM GYGA++ YP R+TA N++ N+GRA+GGF

Sbjct: 305 YSQLTDPDIMLLAGAFGLGMFVNGMLGGYGALMAEAYFTEARATAQNVLFNIGRAVGGFGP 364

Query: 362 VIIGMILDVSNVSMVMLFLASLYIVSFLSML 392
V++G ++ + + LA +Y++ L+ +

Sbjct: 365 VVVGSVVLAYSFQTAIALLAIIYVIDMLATI 395

-2290-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2377> which encodes the amino acid sequence <SEQ ID 2378>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 39
      >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -8.92    Transmembrane  168 - 184 ( 162 - 188)
      INTEGRAL    Likelihood = -5.41    Transmembrane  286 - 302 ( 285 - 306)
      INTEGRAL    Likelihood = -5.15    Transmembrane  372 - 388 ( 362 - 394)
10     INTEGRAL    Likelihood = -3.45    Transmembrane  259 - 275 ( 257 - 276)
      INTEGRAL    Likelihood = -2.87    Transmembrane  311 - 327 ( 306 - 328)
      INTEGRAL    Likelihood = -2.81    Transmembrane   55 - 71 (  51 -  71)
      INTEGRAL    Likelihood = -0.48    Transmembrane  108 - 124 ( 108 - 125)
      INTEGRAL    Likelihood = -0.37    Transmembrane   84 - 100 (  84 - 100)

15     ----- Final Results -----
           bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

20 An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 306/402 (76%), Positives = 354/402 (87%)

Query: 1  MSPLNINKNNWRALIAAIVASGTTDDLNMFLAFMSMTIITDLHLSAAQAGWIGTITNLGM 60
      MS I+++ N RAL+AAI ASGTTDDLNMFLAFMS+I+TDL LS Q GWI TITNLGM
25     Sbjct: 1  MSTLSLDTTNKRALVAAIAASGTTDDLNMFLAFMSMTIITDLHLSAAQAGWIGTITNLGM 60

Query: 61  LVGGLIFGLLADRYNKFVKVFKWTILIFSIATGLVFFTTNLSYLYIMRFTIAGIGVGGEYGI 120
      LVGGL+FGLLADR++KFKVFKWTIL+FS+ATGL++FT +L YLY+MRFTIAGIGVGGEYG+
30     Sbjct: 61  LVGGLLFGLLADRRHKKFKVFKWTILIFSVATGLIYFTQSLPYLYLMRFTIAGIGVGGEYGV 120

Query: 121  AIAIMAGIVPTNKMGRISLNGIAGQVGSISSALLAGWLAPALGWRGLFLFGLLPVLVL 180
      AIAIMAGIVP KMC+SSLNGIAGQ+GSISSALLAGWLAP+LGWRGLFLFGLLP+LV+
35     Sbjct: 121  AIAIMAGIVPPEKMGMSLNGIAGQLGSISSALLAGWLAPSLGWRGLFLFGLLPILLVI 180

Query: 181  WMQFAVDDKIDLDQYNTDADDEPLDISIKALFDTFVLATQSLALMVMTTVQIAGYFGMMN 240
      WM A+DD+ I D Y + ++ I I L F T L Q+LALMVMTTVQIAGYFGMMN
40     Sbjct: 181  WMTLAIDDKIWDHYGQEEECSPQIKINELFKTKSLTAQTLALMVMTTVQIAGYFGMMN 240

Query: 241  WLPTTIQTNLNVSVKNSLWMIATILGMCLGMLVFGQLLDKFGPRLVYGCFLLSSAICVY 300
      WLPTTIQT+LN+SVK+SSLWM+ATI+GMCLGML FGQLLD FGPR+Y FIL+S+ICVY
45     Sbjct: 241  WLPTTIQTSLNLSVKSSSLWVATIVGMCLGMLYFGQLLDGFGPRLIYSLFLLASSICVY 300

Query: 301  LFQFATTPMSMIIGGAVVGGFFVNGMFAGYGAMITRLYPHIRSTANNLILNVGRAIGGFS 360
      LFQFA +M SM+IGGA+VGFFVNGMFAGYGAMITRLYPHIRSTANN+ILNVGRA+GGFS
50     Sbjct: 301  LFQFANSMAVMVIGGAIVGFFVNGMFAGYGAMITRLYPHIRSTANNVILNVGRALGGFS 360

Query: 361  SVIIGMILDVSNVSMVMLFLASLYIVSFLSMLSIKQLKRQKY 402
      SV IG ILD S +SMVM+FLASLY++SF +M SI QLK ++Y
      Sbjct: 361  SVAIGSILDASGISVMVIFLASLYVISFGAMWSIGQLKAERY 402

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2033

55 A DNA sequence (GBSx2144) was identified in *S.agalactiae* <SEQ ID 6285> which encodes the amino acid sequence <SEQ ID 6286>. This protein is predicted to be leucyl-tRNA synthetase (leuS). Analysis of this protein sequence reveals the following:

```

      Possible site: 52
      >>> Seems to have no N-terminal signal sequence

```

-2291-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3481(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

A related GBS nucleic acid sequence <SEQ ID 10241> which encodes amino acid sequence <SEQ ID 10242> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAC00259 GB:AF008220 leucine tRNA synthetase [Bacillus subtilis]
Identities = 569/835 (68%), Positives = 666/835 (79%), Gaps = 42/835 (5%)

15 Query: 10 YNHKEIEPKWQAFWADNHTFKTGTDAASKPKFYALDMFPYPGAGLHVGHPEGYTATDILS 69
+ HKEIE KWQ +W +N TF T + K KFYALDMFPYPGAGLHVGHPEGYTATDILS
Sbjct: 3 FQHKIEIEKKWQTYWLENKTFATLDNNEKQKFYALDMFPYPGAGLHVGHPEGYTATDILS 62

20 Query: 70 RFKRAQGHNVLHPMGWDAFGLPAEQYAMDTGNDPAEFTAEINIANFKRQINALGFSYDWR 129
R KR QG++VLHPMGWDAFGLPAEQYA+DTGNDPA FT +NI NF+RQI ALGFSYDWR
Sbjct: 63 RMKRMQGYDVLHPMGWDAFGLPAEQYALDTGNDPAVFTKQNIIDNFRRIQALGFSYDWR 122

25 Query: 130 EVNTTDPNYYKWTQWIFTKLYEKGLAYEAEVPVNWVEELGTAIANEEVLPDGTSEGGYP 189
E+NTTDP YYKWTQWIF KLYEKGLAY EVPVNW LGT +ANEEV+ DG SERGG+P
Sbjct: 123 EINTTDEYYKWTQWIFLKYEKGLAYVDEVPVNWCPALGTVLANEEVI-DGKSERGGHP 181

30 Query: 190 VVRKPMRQWMLKITAYAEERLLEDLEEDWPESIKDMQRNWIGKSTGANVTFKVKDTDKDF 249
V R+PM+QWMLKITAYAEERLLEDLEEDWPESIKDMQRNWIG+S GA+V F + D F
Sbjct: 182 VERRPMKQWMLKITAYAEERLLEDLEEDWPESIKDMQRNWIGRSEGAHVHFAIDGHDDSF 241

35 Query: 250 TVFTTRPDTLFGATYAVLAPEHALVDAITADQAEAAVEYKQASLKSDDLARTDLAKEKT 309
TVFTTRPDTLFGATY VLAPEHALV+ ITTA+Q EAV Y ++ KSDL RTDLAK KT
Sbjct: 242 TVFTTRPDTLFGATYTVLAPEHALVENITTAEQKEAVEAYIKI QSKSDLERTDLAKTKT 301

40 Query: 310 GVWTGAYAINFPVNGKEIPVWIADYVLASYGTGAIMAVPAHDERDWEFAKQFNLDIIPVLE 369
GV+TGAYAINFPVNG+++P+WIADYVLASYGTGA+MAVP HDERD+EFAK F L + V++
Sbjct: 302 GVFTGAYAINFPVNGEKLPIWIADYVLASYGTGAVMAVPGHDERDFEFAKTFLPVKEVVK 361

45 Query: 370 GGNVEEAAFTEDGLHINSDFLDGLDKAAAIKMWLEAEAGVGNEKVYRLRDWLFQR 429
GGNVEEAA+T DG H+NSDFL+GL K AI K++ WLE G +KVYRLRDWLFQR
Sbjct: 362 GGNVEEAAFTEDGHEVNSDFLNLGHKQEAIEKVIWLEETKNGEKKVYRLRDWLFQR 421

50 Query: 430 YWGEPIPIIHWDGTSTAVPESELPLVLPVTKDIRPSGTGESPLANLTDWLEVT-REDGV 488
YWGEPIP+IHWDGTSTAVPE ELPL+LP T +I+PSGTGESPLAN+ +W+EVT E G
Sbjct: 422 YWGEPIPVIHWEDGTSTAVPEBELPLILPKTDEIKPSGTGESPLANIKWVEVTDPETGK 481

55 Query: 489 KGRREINTMPQWAGSSWYLYRIDPHNTEKLADDELLKQWLPVDIYVGAHAHLHLLYA 548
KGRREINTMPQWAGS WY+LYRIDPHN ++LA E L++WLPVD+Y+GGAHAHLHLLYA
Sbjct: 482 KGRREINTMPQWAGSCWYFLRYIDPHNPDQLASPEKLEKWLVDYIGGAHAHLHLLYA 541

60 Query: 549 RFWHKVLVDLGVVPTKEPFQKLFNQGMILGTSYRDSRGALVATDKVEKRDGSFFHVETGE 608
RFWHK LYD+GVVPTKEPFQKL+NQGMILG E E
Sbjct: 542 RFWHKFLYDIGVVPTKEPFQKLYNQGMILG-----ENNE 575

65 Query: 609 ELEQAPAKMSKSLKNVNPDDVVEQYGADTLRVYEMFMGPLDASIAWSEEGLEGSRKFLD 668
KMSKS NVVNPDD++V +GADTLR+YEMFMGPLDASIAWSE GL+G+R+FLD
Sbjct: 576 -----KMSKSGNVNPDEIVASHGADTLRLYEMFMGPLDASIAWSEGLDGARRFLD 628

Query: 669 RVYRLI-----TTKEITEENSGALDKVYNETVKAVTEQVDQMKFNIAIQLMVFNAN 722
RV+RL +I E L++VY+ETV VT+ + ++FNT I+QLMVF+N A
Sbjct: 629 RVWRLFIEDSGELNGKIVEGAGETLERYVHETVMKVTVDHYEGLRFNTGISQLMVFINEAY 688

Query: 723 KEDKLFSDYAKGFVQLIAPFAPHLGEELWQVLTASGQSISYVPWPSYDESKLVENEIEIV 782
K +L +Y +GFV+L++P APHL EELW+ L SG +I+Y WP YDE+KLV++E+IEIV
Sbjct: 689 KATELPKEYMBGFVKLLSPVAPHLAEELWEKLGHS-GTAYEAWPVYDETKLVDEVEIV 747

Query: 783 VQIKGKVKAKLVVAKDLSREELQDLALANEKVQABIAGKDIIKVIAPNKLNVIV 837
VQ+ GKVKAKL V D ++E+L+ LA A+EKV+ ++ GK I K+IAPN KLNVIV

-2292-

Sbjct: 748 VQLNGKVKAKLQVPADATKEQLEQLAQADEKVKKEQLEBKTKIRKIIAVPGKLVNIV 802

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6287> which encodes the amino acid sequence <SEQ ID 6288>. Analysis of this protein sequence reveals the following:

5 Possible site: 46
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.4358(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 815/833 (97%), Positives = 827/833 (98%)

15 Query: 7 MTFYNHKEIEPKWQAFWADNHTFKTGT DASKPKFYALDMFPYPGAGLHVGHPEGYTATD 66
 MTFY+H IEPKWQAFWADNHTFKTGT DASKPKFYALDMFPYPGAGLHVGHPEGYTATD
 Sbjct: 1 MTFYDHTAIEPKWQAFWADNHTFKTGT DASKPKFYALDMFPYPGAGLHVGHPEGYTATD 60

20 Query: 67 ILSRFKRAQGHNVLHPMGWDAFGLPAEQYAMDTGNDPAEFTAENIANFKRQINALGFSYD 126
 ILSRFKRAQGHN+LHPMGWDAFGLPAEQYAMDTGNDPAEFTAENIANFKRQINALGFSYD
 Sbjct: 61 ILSRFKRAQGHNVLHPMGWDAFGLPAEQYAMDTGNDPAEFTAENIANFKRQINALGFSYD 120

25 Query: 127 WDEVNTTDPNYYKWTQWIFTKLYEKGLAYEAEVFNWVEELGTAIANEEVLPDGT SERG 186
 WDEVNTTDPNYYKWTQWIFTKLYEKGLAYEAEVFNWVEELGTAIANEEVLPDGT SERG
 Sbjct: 121 WDEVNTTDPNYYKWTQWIFTKLYEKGLAYEAEVFNWVEELGTAIANEEVLPDGT SERG 180

30 Query: 187 GYPVVRKPMRQWMLKITAYAERLLEDLEEVDWPESIKDMQRNWIGKSTGANVTFKVKDITD 246
 GYPVVRKPMRQWMLKITAYAERLLEDLEEVDWPESIKDMQRNWIGKSTGANVTFKVKDITD
 Sbjct: 181 GYPVVRKPMRQWMLKITAYAERLLEDLEEVDWPESIKDMQRNWIGKSTGANVTFKVKDITD 240

35 Query: 247 KDFTVFTTRPDTLFGATYAVLAPPEHALVDAITTADQAEAVA EYKRQASLKS DLARTDLAK 306
 KDFTVFTTRPDTLFGATYAVLAPPEHALVDAITTADQAEAVA+YKRQASLKS DLARTDLAK
 Sbjct: 241 KDFTVFTTRPDTLFGATYAVLAPPEHALVDAITTADQAEAVAKYKRQASLKS DLARTDLAK 300

40 Query: 307 EKTGVWVTGAYAINPVNGKEIPVWIADYVLASYGTGAIMAVPAHDERDWEFAKQFNLDIIP 366
 EKTGVWVTGAYAINPVNG E+PVWIADYVLASYGTGAIMAVPAHDERDWEFAKQFNLDIIP
 Sbjct: 301 EKTGVWVTGAYAINPVNGNEMPVWIADYVLASYGTGAIMAVPAHDERDWEFAKQFNLDIIP 360

45 Query: 367 VLEGGNVEEAAFTEDGLHINSDFLDGLDKAAAIKAMVEWLEAEGVGNEKVITYRLRDWLFS 426
 VLEGGNVEEAAFTEDGLHINS FLDGLDKA+AIKAMVEWLEAEGVGNEKVITYRLRDWLFS
 Sbjct: 361 VLEGGNVEEAAFTEDGLHINSDFLDGLDKASAIKAMVEWLEAEGVGNEKVITYRLRDWLFS 420

50 Query: 427 RQRYWGEPPIIIHWEDGTSTAVPESELPLVLPVTKDIRPSGTGESPLANLTDWLEVTRED 486
 RQRYWGEPPIIIHWEDGTSTAVPESELPLVLPVTKDIRPSGTGESPLAN+TDWLEVTRED
 Sbjct: 421 RQRYWGEPPIIIHWEDGTSTAVPESELPLVLPVTKDIRPSGTGESPLANVTDWLEVTRED 480

55 Query: 487 GVKGRRETNTMPQWAGSSWYYLRYIDPHNTEKLADDELLKQWLPVDIYVGGAEHAVLHLL 546
 GVKGRRETNTMPQWAGSSWYYLRYIDPHNTEKLADDELLKQWLPVDIYVGGAEHAVLHLL
 Sbjct: 481 GVKGRRETNTMPQWAGSSWYYLRYIDPHNTEKLADDELLKQWLPVDIYVGGAEHAVLHLL 540

60 Query: 547 YARFWHKVLYDLGVVPTKEPFQKLFNQGMILGTSYRDSRGALVATDKVEKRDGSFFHVET 606
 YARFWHKVLYDLGVVPTKEPFQKLFNQGMILGTSYRDSRGALVATDKVEKRDGSFFHVET
 Sbjct: 541 YARFWHKVLYDLGVVPTKEPFQKLFNQGMILGTSYRDSRGALVATDKVEKRDGSFFHVET 600

65 Query: 607 GEELEQAPAKMSKSLKNVNPDDVVEQYGADTLRVYEMFMGPLDASIAWSEEGLEGSRKF 666
 GEELEQAPAKMSKSLKNVNPDDVVEQYGADTLRVYEMFMGPLDASIAWSEEGLEGSRKF
 Sbjct: 601 GEELEQAPAKMSKSLKNVNPDDVVEQYGADTLRVYEMFMGPLDASIAWSEEGLEGSRKF 660

60 Query: 667 LDRVYRLITTKEITEENSGALDKVYNETVKAVTEQVDQMKFNATAIQLMVFVNAANKEDK 726
 LDRVYRLITTKEITEENSGALDKVYNETVKAVTEQVDQMKFNATAIQLMVFVNAANKEDK
 Sbjct: 661 LDRVYRLITTKEITEENSGALDKVYNETVKAVTEQVDQMKFNATAIQLMVFVNAANKEDK 720

65 Query: 727 LFSDYAKGFVQLIAPFAPHLGEELWQVLTASGQSISYVPWPSYDESKLVENEIEIVVQIK 786
 LFSDYAKGFVQLIAPFAPHLGEELWQ LTASG+SISYVPWPSYDESKLVEN++EIVVQIK

-2293-

Sbjct: 721 LFSDYAKGFVQLIAPFAPHLGELWQALTASGESISYVPWPSYDESKLVENDVEIVVQIK 780

Query: 787 GKVKAKLVVAKDLSREELQDLALANEKVQAEIAGKDIIKVIAPNKLNVIVVK 839
GKVKAKLVVAKDLSREELQ++ALANEKVQAEIAGKDIIKVIAPNKLNVIV+K

5 Sbjct: 781 GKVKAKLVVAKDLSREELQAEVALANEKVQAEIAGKDIIKVIAPNKLNVIVIK 833

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2034

10 A DNA sequence (GBSx2145) was identified in *S.agalactiae* <SEQ ID 6289> which encodes the amino acid sequence <SEQ ID 6290>. This protein is predicted to be KIAA1074 protein. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have an uncleavable N-term signal seq

15

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

20

A related GBS nucleic acid sequence <SEQ ID 8961> which encodes amino acid sequence <SEQ ID 8962> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 7

SRCFLG: 0

25

McG: Length of UR: 19

Peak Value of UR: 2.86

Net Charge of CR: 4

McG: Discrim Score: 10.27

GvH: Signal Score (-7.5): -3.61

30

Possible site: 31

>>> Seems to have an uncleavable N-term signal seq

Amino Acid Composition: calculated from 1

ALOM program count: 0 value: 2.12 threshold: 0.0

35

PERIPHERAL Likelihood = 2.12 7

modified ALOM score: -0.92

*** Reasoning Step: 3

----- Final Results -----

40

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

45 No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8962 (GBS117) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 8; MW 22.5kDa).

GBS117-His was purified as shown in Figure 200, lane 7.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

50

-2294-

Example 2035

A DNA sequence (GBSx2146) was identified in *S.agalactiae* <SEQ ID 6291> which encodes the amino acid sequence <SEQ ID 6292>. This protein is predicted to be YirC (resE). Analysis of this protein sequence reveals the following:

```

5      Possible site: 28
      >>> Seems to have an uncleavable N-term signal seq
          INTEGRAL      Likelihood = -10.88      Transmembrane 177 - 193 ( 173 - 196)
          INTEGRAL      Likelihood = -4.09      Transmembrane 10 - 26 ( 5 - 29)

10     ----- Final Results -----
          bacterial membrane --- Certainty=0.5352(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15     The protein has homology with the following sequences in the GENPEPT database.

      >GP:CAB15292 GB:Z99120 similar to two-component sensor histidine
          kinase [YvqA] [Bacillus subtilis]
          Identities = 108/379 (28%), Positives = 193/379 (50%), Gaps = 33/379 (8%)

20     Query: 92  DNHHKESHDIIRYLTQKRLWQISKEKGMFVTIKKKTYVMTKDYSGILVDGSIKKVPKA 151
          +N + S + L+ + ++ K D KKK Y + D +G V IKK
      Sbjct: 86  ENEEASSDKDLSILSSSFIHKVYKLADKQ--EAKKKRY---SADVNGEKVFFVIKKGLSV 140

      Query: 152  QSQLFHVINF-----DITYTQHLITKINHFLIVILVLTYPMLFIMRKTFTGIRESIQ 205
          Q +++++ D+ YT L ++ + V+++L++IP +++ + + +
25     Sbjct: 141  NGQSAMMLSVALDSYRDDLAYT--LFKQLLFIIAVVILLSWIPAIWLAKY----LSRPLV 194

      Query: 206  SVQTYISSLWKNQGNHQSSQKEIVFSDFDPLLESQEMANRIYQAEESQRNFFQNASHEL 265
          S + ++ + ++ + K + L +EM ++ Q +E++R QN SH+L
30     Sbjct: 195  SFEKHVKRI--SRQDWDDPVKVDKDEIGKLGHTIEEMRQKLVQKDETERTLQNISHDL 252

      Query: 266  RTPMLSIQGYTEGVQEGII---DAELAHSVILQESKMKQLVDDIILLKLD--SNLSDQ 320
          +TP+M I+GYT+ +++GI D E VI E+ K+++ + D++ L+KLD + Q
35     Sbjct: 253  KTPVMVIRGYTQSIKDGIFPKGDLENTVDVIECEALKLEKKIKDLLYLTCLDYLAQKQVQ 312

      Query: 321  KDEFSLNELNLSIIAYFKPLANKQKISITYRPDKHEKLLK-GNEELIQRAINNILSNALR 379
          D FS+ E+ +I K A K+ +++ D E +L G+ E + + NIL N +R
      Sbjct: 313  HDMFSIVEVTEEVIERLK-WARKE---LSWEIDVEEDILMPGDPEQWNKLENILENQIR 368

40     Query: 380  YAVSHIEISYT----NQKLTIISNDGPAISKEDLFYIFDRFYKGHGGQTGIGLAMTKKIIK 435
          YA + IEIS N +TI NDGP I E L +++ F KG G+ GIGL++ K I+
      Sbjct: 369  YAETKIEISMKQDDRNIVITIKNDGPHIEDEMLSSLYEPFNKGGKGEFGIGLSIVKRILT 428

      Query: 436  QHNGNIIAESDSTSTTFTI 454
          H +I E+D T ++ I
45     Sbjct: 429  LHKASISIENDKTGVSYRI 447
  
```

There is also homology to SEQ ID 1178.

SEQ ID 6292 (GBS279) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 7; MW 54.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 6; MW 79.4kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2295-

Example 2036

A DNA sequence (GBSx2147) was identified in *S.agalactiae* <SEQ ID 6293> which encodes the amino acid sequence <SEQ ID 6294>. This protein is predicted to be two-component response regulator (mtrA). Analysis of this protein sequence reveals the following:

```

5   Possible site: 37
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.1706(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10239> which encodes amino acid sequence <SEQ ID 10240> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB05663 GB:AP001513 two-component response regulator [Bacillus halodurans]
Identities = 87/220 (39%), Positives = 124/220 (55%), Gaps = 4/220 (1%)

Query: 11 IYFADDEKNIRDLVVPFLEHDGFTVRAFETGDLLEAYKNQKPDVLVILDIMPGTNGLDV 70
20      I DDE ++R+LV +L +GF V ETGD ++ + + DLV+LD+MM +G
Sbjct: 7 ILIVDDELDELRELVTSYLRKEGFAVYTAETGDEAIKRLEQEPMDLVVLDVMMDEMMDGFTA 66

Query: 71 MKSIRQYDNIPIMIILTARDSIDVDFITAFNLGTDYFTKPFSPIKLSLHVKALFKRLDEKA 130
      K IR + IPIIMLTAR + D + +G DDY KPFSP +L ++ +R
25 Sbjct: 67 CKEIRAFSQIPTIIMLTARGGEDDKVMGLQIGADDYIVKPFSPRELVARIEVALRRTQGIQ 126

Query: 131 IKNDTQYQFLDLTLDTEKRIALLSNEEMPLTKTEFDFLLVLEKPEAFSRETLLNRWIG 190
      +DT Y+F +L + R ++ +E+ LTK E+D L+ L+E F+RE L +R+WG
30 Sbjct: 127 QVDDTGYRFNELRIQPSGRKVFVNGQEBISLTKKEYDLLVFLEHGRVFTREHLHDRLWG 186

Query: 191 FDDIES--RAVDDTIKRLRKKFKQYHSQVSIKTVWGYGFK 228
      D + R VD IK LR K K + IKT VWG G+K
Sbjct: 187 MDTQQGTLRTVDTHIKTLRLKLP--ADRFIKTVWGVGYK 224

```

35 There is also homology to SEQ ID 3260.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2037

40 A DNA sequence (GBSx2148) was identified in *S.agalactiae* <SEQ ID 6295> which encodes the amino acid sequence <SEQ ID 6296>. Analysis of this protein sequence reveals the following:

```

Possible site: 55
>>> Seems to have a cleavable N-term signal seq.
      INTEGRAL Likelihood = -2.18 Transmembrane 1568 -1584 (1568 -1585)
      INTEGRAL Likelihood = -0.16 Transmembrane 338 - 354 ( 338 - 354)
45 ----- Final Results -----
      bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10237> which encodes amino acid sequence <SEQ ID 10238> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-2296-

>GP:AAG09771 GB:AF243528 cell envelope proteinase [Streptococcus thermophilus]
Identities = 797/1594 (50%), Positives = 1056/1594 (66%), Gaps = 39/1594 (2%)

5 Query: 21 MNTKQRFKIRKYKLGAVSVLLGLTFLGGITNVAADSVINKPSDIAVEQQVKDSPTS-IA 79
M K+ FS+RKYK+G VSVLLG +F G +VAAD + + + VE V D+ S A
Sbjct: 1 MKKKETFSLRKYKIGTVSVLLGAVFLFAGAPSVAADELTSLV-ETKVEATVPDAIVSESA 59

10 Query: 80 NETPTNN--TSSALASTAQDNLVTKANNSTPTETQPVASHSQATETFSFVANQPVESTQE 137
+E+P +++ +T+ D T ++ + S + ET P P S ++
Sbjct: 60 SESPPVVEELVDTSVEATST'DVTTTDNEEETPGSEALENSANTEVETTQPAVETPAISEKK 119

15 Query: 138 VSKTPLTKQNLAVKSTPAISKET--PQNIDSNKIITVPKVMNTGYKGEGETVVAIIDSGLD 195
V + K ++A ++T ++E PQNIDSN IITVPKVV +GYKGEGETVVAIIDSGLD
Sbjct: 120 VEBEE--KLSVADETTAITNQEEAKPQNIDSNITITVPKVMWVSGYKGEGETVVAIIDSGLD 177

20 Query: 196 INHDALQLNDSTKAKYQNEQQMNAAKAKAGINYKWNKVI FGHNYVDVNTTELKEVKST 255
++HD L ++D + AKY++E+++ AAK AGI YG+W+N+KV+FG+NYVDVNT LKE
Sbjct: 178 VDHDVLHISDLSTAKYKSEKEIEAAKEAAGITYGEWFNDKVVFGYNYVDVNTVLKEEDKR 237

25 Query: 256 SHGMHVTSIATANPSKKDNTNELIYGVAPEAQVMFMRVFSDEKRGITGPALYVKAIEDAVKL 315
SHGMHVTSIAT NP++ +L+YGVAPEAQVMFMRVFS D K TG ALYVKAIEDAVKL
Sbjct: 238 SHGMHVTSIATGNPTQPVAGQLMYGVAPEAQVMFMRVFS DLKATTGAALYVKAIEDAVKL 297

30 Query: 316 GADSLNLSLGGANGSLVNADRLIKALEMARLAGVSVVIAAGNDGTFGSGASKPSALYPD 375
GADSLNLSLGGANGS+VN ++ + A+E AR AGVSVVIAAGNDGTFGSG S PSA YPD
Sbjct: 298 GADSLNLSLGGANGSVVMNENVTAAIEAARRAGVSVVIAAGNDGTFGSGHNSPSADYPD 357

35 Query: 376 YGLVGSPTAREAISVASYNNTTLVNKVFNIIGLENNRNLNGLAAYADPKVSDKTFEVG 435
YGLVG+PSTA +AISVASYNNTT+ +KV NIIGLENN +LN G +++ +P+ S FE+G
Sbjct: 358 YGLVGAPSTAHDALISVASYNNTTVGSKVINIIGLENNADLNKSSFDNPEKSPVPFEIG 417

40 Query: 436 KQYDYVFGKGNNDYKDKTLNGKIALIERGDITFTKKVNVAINHGAVGAIIFNNKAGEA 495
K+Y+YV+ G G +D+ L GK+ALI+RG ITF++K+ NA GAVG +IFN++ GEA
Sbjct: 418 KEYEYVYAGIGQASDFDGLDLTGKLALIKRGITITFSEKIANATAAGAVGVVIFNSRPGEA 477

45 Query: 496 NLTMSLDPEASAIPAIFTQKEFGDVLAKNNYKIVFNNIKNKQANPNAGVLSDFSSWGLTA 555
N++M LD A AIP++F EFG+ LA N+YKI FNN + + NP AG+LSDFSSWGL+A
Sbjct: 478 NVSMQLDDTAIAIPSVFIPLEFGEALAANSYKIAFNNETDIRNPGEGLLSDFSSWGLSA 537

50 Query: 556 DGQLKPDLSAPGGSIYAAINDNEYDMMSGTSMASPHVAGATALVKQYLLKEHPELKKGDI 615
DG+LKPDL+APGG+IYAAINDN+Y M GTSMASPHVAGA LVKQYLL +P +I
Sbjct: 538 DGELKPDLAAPGCAIYAAINDNDYANMQGTSMASPHVAGAAVLVKQYLLATYPTKSPQEI 597

55 Query: 616 ERTVKYLLMSTAKAHLNKDTGAYTSPRQQAGIIDVAAVQTGLYLTGGENNYGSVTLGN 675
E VK+LLMSTAKAH+NK+T AYTSPRQQAGIID AAA+ TGLYLT GE+ YGS+TLGN
Sbjct: 598 EALVKHLLMSTAKAHVNKETTAYTSPRQQAGIIDTAAAI+TGLYLT-GEDGYGSITLGN 656

60 Query: 676 IKDKISFDVTVHNINKVAKDLHYTTYLTNDQVKDGFVTLAPQQLGTFGTIRIEPGQTQ 735
++D SF VT+HNI K L+Y+T L TD + L + + + + + +
Sbjct: 657 VEDTFSFTVTLHNITNEDKTLNYSTQLTDTAQKRIDLHGSTSISRDSWRKVTVKANSST 716

65 Query: 736 TITIDIDVSKYHDMKKVMPNGYFLEGYVRFTDPVDGGEVLSIPYVGFKEFQNLVLEK 795
T+TI++D S + + L +M NGY+LEG+VRFTD D G+++SIPYVGF+GEFQNL VLE+
Sbjct: 717 TVTINVDASSFABEELTGLMKNGYILEGFVRFTDVADGDIVSIPYVGFGRGEFQNLAVLEE 776

70 Query: 796 SIYKLVANKEKGFYFQP--KQTNEVPGSEDTALMTTSSEPIYSTDGTSPILKALGSYK 853
IY L+A+ + GFYF+P Q N V S YT L+T S+E IYSTD S +K LG++K
Sbjct: 777 PIYNLIADGKGGFYFEPVTAQPNVTVDISHHYTGLVTGTSLIYSTDKRSDSAIKTLGTFK 836

75 Query: 854 SIDGKWILQLDQKGQPHLAISPNDQNDQDAVAVKGVFLRNFNRLRAKVYRADDVNLQKPL 913
+ G ++L+LD+ G+PHLAISPND NQD++ KGVFLRN+ +L A VY ADD PL
Sbjct: 837 NKAGYFVLELDESQKPHLAISPNGDDNQDSL VFKGVFLRNYTDLVASVYAADDTERTNPL 896

80 Query: 914 WVSAPQAGDKNYISGNTENPKSTFLYDTEWKGTTTDGIPLEDGKYKYVLTYYSVDPGSKP 973
W S PQ+GDKN YSGN +NPKS+ +Y TEW GT +DG L DGKY+YVLTY S VPG+
Sbjct: 897 WESQPPQSGDKNIYSGNPKNPKSSIIYPTWNGTSDGNALADGKYQYVLTYSSKVPGAHV 956

85 Query: 974 QQNVFDITLDRQAPTLLTATYDKDRRIFKARPAVEHGESGIFREQVYFKKDKDGHYNSV 1033

-2297-

Q M+FD+ +DR++P +TTATYD+ F RPA+E GESG++REQVFYL D G ++
 Sbjct: 957 QTMIFDVIIDRESFVITATYDETNTFTNPRPAIEKGESGLYREQVFYLVADASG-VTTI 1015

Query: 1034 LRQQGEDGILVEDNKVFIKQEKDGSFILPKEVNDFSHVYTVEDYAGNLVSAKLEDLINI 1093
 + V DNKVF+ Q DGSF LP ++ D S YTVEDYAGN+ K+E+LI+I
 Sbjct: 1016 PSLKNGDVTVDNKFVFAQNDGDSFTLPLDLADISKFYTVEDYAGNISYEKVENLISI 1075

Query: 1094 GNKNGLVNVKVFSPELNSNVDIDFSYSVKDDKGNIIKK-QHHGKDLNLLKLPFGTYTDFL 1152
 GN+ GLV V + + NS V I FSYSV D+ G I+ + + D ++LKLPGFTYTFDL
 Sbjct: 1076 GNEKGLVTNILDKDTNSPVPILFSYSVTDETGKIVAEPLRYAGDTSVLKLPFGTYTDFL 1135

Query: 1153 FLYDEERANLISPKSVTVTISEKDSLKDVLKVNLLKKAALLVEFDKLLPKGATVQLVTK 1212
 FLYD E ++L VTI E +S +V F V L KA LL++ D LLP G+T+QLVT
 Sbjct: 1136 FLYDTEWSSLAGETKAVVTILEDNSTAEVNFYVTLKDKANLLIDIDALLPSGSTIQLVTA 1195

Query: 1213 TINTVVDLPKATYSPTDYGKNIPVG DYRLNVTLPSGYSTLENLDDLVSVKEDQVNLTKLT 1272
 + LP A YS TDYGK +PVG Y + TLP GY LE LD V+V +Q N+ KLT
 Sbjct: 1196 DGQAIQLPNAKYSKTDYGKFPVPGTYTILPTLPEGYEFLELD---VAVLANQSNVKKLT 1252

Query: 1273 LINKAPLINALAEQTDIITQPVFYNAGTHLKNNYLANLEKAQTLIKNRVEQTSIDNAIAA 1332
 LINK L +AE + +YNA L+ Y LE A + N+ Q +D+A+A+
 Sbjct: 1253 LINKVALKELIAELAGLEETARYYNASPELQTAYAKALEDANAVYANKHNQAQVDSALAS 1312

Query: 1333 LRESRQALNGKETDTSLLAKAILAETEIKGNYQFVNASPLSQSTYINQVQLAKNLLQKPN 1392
 L +R+ LNG+ TD L + T + N+ + NA Q Y V+ A+ +L + N
 Sbjct: 1313 LVAAREQLNGQATDKEKLIAEVSNYTPQANFIYNAENTKQIAYDTAVRSAQLVLNQEN 1372

Query: 1393 VTQSEVDKALENLDIAKNQNLNGHETDYSGLHHMIKANVLKQTSSKYQNASQFAKENYNN 1452
 VTQ+ V++AL +L AK L+G +TD S L + ++VLK T +KY NAS+ K+ Y+
 Sbjct: 1373 VTQAVVNQALADLLAAKANLDGQKTDISALRSVSVSSVLKATDAKYLNASENVKQAYDQ 1432

Query: 1453 LIKKAELLLSNRQATQAQVEELLNQIKATEQEELDG----RDRVSSAENYSQSLNDNDSLN 1508
 ++ A+ +L + A+QA V++ L + + + ELDG + N + D ++
 Sbjct: 1433 AVEAAKAILVDESASQASVDQALAVLTSQAQAEIDGVATSTNDAKEPANTATDKKDEGTVT 1492

Query: 1509 TTPIN-----PP-----NQPALIFKKGMTKESEVAQKRVLGVTSTQTDNQVKTKNL 1555
 PI+ PP N I +K + + + L + + NQ+ + +L
 Sbjct: 1493 PPPIDSEIVDVQAPPVKDTGNSEHVPIGQK-PNPQPTLPRPVTLQASLSPNQEKQVTQL 1551

Query: 1556 PKTGESTPKITYTILFSLMLGLATIKLSIKR 1589
 P TGE+ K L ++GL T+ L SI+R
 Sbjct: 1552 PNTGENDTK---YYLVPGVIIGLGT-LVSIRR 1580

A related GBS gene <SEQ ID 8963> and protein <SEQ ID 8964> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 SRCFLG: 0
 McG: Length of UR: 1
 Peak Value of UR: 2.55
 Net Charge of CR: 4
 McG: Discrim Score: 2.60
 GvH: Signal Score (-7.5): -0.78
 Possible site: 35
 >>> Seems to have a cleavable N-term signal seq.
 Amino Acid Composition: calculated from 36
 ALOM program count: 1 value: -0.16 threshold: 0.0
 INTEGRAL Likelihood = -0.16 Transmembrane 318 - 334 (318 - 334)
 PERIPHERAL Likelihood = 2.54 1161
 modified ALOM score: 0.53
 icml HYPID: 7 CFP: 0.106

*** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2298-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 1535-1539

5 The protein has homology with the following sequences in the databases:

50.5/67.5% over 1583aa

Streptococcus

s thermophilus

GP|9963932| cell envelope proteinase Insert characterized

10

ORF01603(361 - 5070 of 5370)

GP|9963932|gb|AAG09771.1|AF243528_1|AF243528(1 - 1584 of 1585) cell envelope proteinase {Streptococcus thermophilus}

%Match = 41.2

15

%Identity = 50.4 %Similarity = 67.4

Matches = 794 Mismatches = 498 Conservative Sub.s = 267

20

255	285	315	345	375	405	435	465
KNALGTVLNL	PQNNL**KFRKL*	KILIFVVLIVFV	IIMLQKEIFM	NTKQRF	SIRKYKLGAVSV	LGLTFL	FGGITNVAA
					:	:	:
				MKKKET	PSLRKYKIGTVSV	LGLGAVFL	FAGAPSVAA
				10	20	30	

25

495	525	552	576	606	636	666	696
DSVINKPSDIAVE	QQVKDSPTS-IA	NETPT--NNTSS	ALASTAQDNL	VTKANN	SPPTETQPVA	ESHSQATET	FSPVANQPV
:	:	:	:	:	:	:	:
DE-LTSLVETK	VEATVPDAIV	SESASESPV	VEELVDTS	VEATSDVT	TTDNEEET	PGSEALENS	ANTEVETTPA
50	60	70	80	90	100	110	

30

726	756	780	810	840	870	900	930
ESTQEVSKT	PLTKQNLAVK	STPAISKE--	TPQNIDS	NKIITVPK	VWNTGYK	GEGTVVAI	IDSGLDINH
:	:	:	:	:	:	:	:
ISEKKV	EEEE--	KLSVADETTA	ITNQEEAKP	QNIDSNTI	ITVPKW	VWYSGYK	GEGTVVAI
130	140	150	160	170	180	190	

35

960	990	1020	1050	1080	1110	1140	1170
YQNEQQMNA	AKAKAGIN	YGKWNKVI	FGHNYVDV	NTELKEVK	STSHGMH	VTSIATAN	PSKKDTN
:	:	:	:	:	:	:	:
YKSEKEI	EAAKEA	AGITYGE	WFNDKV	VFGYNY	VDVNTVL	KEEDKR	SHGMH
210	220	230	240	250	260	270	

40

1200	1230	1260	1290	1320	1350	1380	1410
RVFSDEK	RCTGPAL	VKAIEDA	VKLGADS	INLSLGG	ANGSLVN	ADDRLI	KALEMAR
:	:	:	:	:	:	:	:
RVFSDL	KATTGA	ALYKAIE	DAVKLGADS	INLSLGG	ANGSVVN	MNENVT	AATEAARR
290	300	310	320	330	340	350	

45

1440	1470	1500	1530	1560	1590	1620	1650
ALYPDYGL	VGSPSTARE	ASVSYNNT	TLVNKVFNI	IGLENNR	NLNGLAAY	ADPKVSD	KTFEVGK
:	:	:	:	:	:	:	:
ADYPDYGL	VGAPSTA	HAISVSYN	NTTVGSK	VINIIGLENN	ADLN	YGKSSFDN	PEKSPVP
370	380	390	400	410	420	430	

50

1680	1710	1740	1770	1800	1830	1860	1890
YKDKTLN	GKIALIER	GDTFTK	KVVNAI	NHGAVGAI	IFNNKAGE	ANLTMSLD	PEASAI
:	:	:	:	:	:	:	:
FDGLDL	TGKLLAL	IKRGTIT	FSEKIAN	ATAAGAV	GVVIFN	SRPGEAN	VSMQLDD
450	460	470	480	490	500	510	

55

1920	1950	1980	2010	2040	2070	2100	2130
NNIKNKQ	ANPNAGV	LSDFSSW	GLTADGQ	LKPDLSA	PGGSIYAA	INDNEYD	MMSGTSM
:	:	:	:	:	:	:	:
NNETD	IRENPEA	GLLSDF	SSWGLS	ADGELK	PDLAAPG	GAIAA	INDNDYAN
530	540	550	560	570	580	590	

60

2160	2190	2220	2250	2280	2310	2340	2370
KKGDI	ERTV	KYLLM	STAKAHL	NKDTG	AYTSR	QQGAGI	IDVAAAV
QGTGLYLTGGENNYGSVTLGNIKDKISFDVTVHNIN							

65